STEREOPSIS AND THE COMBINATION OF SURFACE CUES

Final Report for Contract N00014-87-K-0321

Kent A. Stevens
Department of Computer Science

Jacob Beck
Department of Psychology

University of Oregon
Eugene, OR 97403

July 10, 1990

Prepared for

Project Manager: Dr. Harold Hawkins

The Office of Naval Research
Perceptual Science Program
ONR Code 1142PS
Arlington, VA 22217

CONTENTS

PART I Binocular Depth and the Construction of Visual Surfaces (Stevens)

PART II Apparent Surface Orientation and Orthographic Projection (Beck)

ABSTRACT

This report describes research regarding the integration of spatial information. Part I (Stevens) reports work that addresses questions of integration, including the form of the spatial information provided by human stereopsis towards the perception of visual surfaces and the strategies by which this information is reconciled with monocular 3D information. Part II (Beck) concerns how surface orientation and distance are perceived in wire-frame figures that are projected orthographically.

PART I

BINOCULAR DEPTH AND THE CONSTRUCTION OF VISUAL SURFACES

Final Report ONR Grant N000-K-0321

Kent A. Stevens

This report summarizes research performed in collaboration with Allen Brookes, whose Ph.D. dissertation, supported by the ONR, was completed in 1988. With an extension to the grant provided by the ONR, Brookes continued as Research Associate.

1. Introduction

To place this effort in context, we must first stress that the fundamental primitives of form perception are as yet unknown. Intuition has suggested to many investigators that depth (the impression of surface relief, or of local variations in distance across a surface and between surfaces) constitutes the fundamental basis on which surfaces are described within the visual system. The primitives that have been offered for the internal representation of surfaces include the depth and surface normal at individual surface patches, and the loci where the surface is discontinuous in either depth or orientation. What has made these choices seemingly tractable and plausible is the fact that these quantities correspond to what appears to be deliverable by various putative visual modules (the "shape from" modules such as orientation from shading, depth from motion, and so forth). The mathematical interconvertability of these quantities provides further support for this approach, since there are attractive lattice computations that can operate on local neighborhoods of such quantities in order to fit smooth surfaces through and between sample points. However, as we will discuss, our recent research has established the primacy of curvature and discontinuity features (at least for stereopsis, and likely for motion, based on observations by other investigators), and the secondary or subsequent nature of depth.

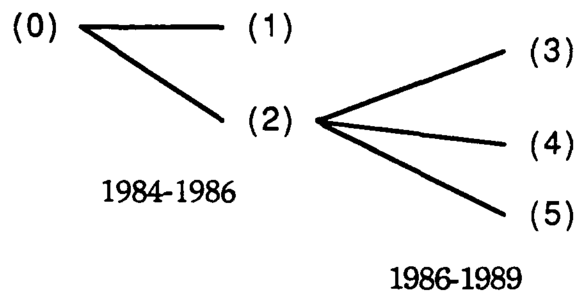
Our observation that stereo depth is a derived, or reconstructed, quantity is not strictly at odds with the notion of an internal representation of surfaces in terms of depth and other scalar quantities. However, our concurrent investigation of the integration of monocular (primarily perspective and foreshortening) cues with stereopsis has revealed cases for which the apparent depth is difficult to explain in terms of existing computational models. Measures of the *geometrical compatibility* of the surface descriptors provided by different sources seem to govern the end percept, and moreover, the compatibility "rules", if we can eventually characterize them as rules, seem to involve some degree of scrutiny. It is this nature of surface perception which we will address. Computationally, the questions concern the introduction of new primitive descriptors for surface events beyond the simple notions of scalar quantities and discontinuity loci, the question of how to impose intervention on the local behavior of the

network, and how to "read" the stable solutions of the network. There are many facets of human behavior that we believe are central to the construction of surface descriptions that have yet to be adequately captured. Insight into that behavior will come from further psychophysical experiments motivated by these computational notions.

The first ONR period (1984-1986) examined interactions among individual monocular cues, specifically surface contours, texture gradients, and shading. The fundamental computational questions then concerned the extent to which different representations, say of surface orientation and of depth are coupled, and the hypothesis that there might be higher-order geometric features involved. An early experimental result, reported in (Stevens & Brookes, 1987) and indicated by (1) in figure 1, demonstrated that the binocular depth of a probe point could be made commensurate with the apparent depth across a purely monocular rendering of a slanted surface, including the difficult case of a surface rendered in orthographic projection. The significance of this result was that depth and slant information are not only intimately related mathematically, but the visual system can readily make them commensurate. This somewhat unexpected result put us on our guard against naive interpretation of the results of psychophysical depth probing experiments. The interpretation of experimental results is complicated by the difficulty in attributing a given judgment to the accessing of a particular internal representation. This difficulty made us reconsider what hypotheses could be tested by direct depth probing.

In 1985, we (Brookes and Stevens) turned to investigate the strategy of the integration process, rather than the magnitude of the percept under different experimental conditions. Richards' intriguing suggestion was that we determine whether monocular perspective and stereo cues were mutually constraining, along the lines of motion and stereo (Richards, 1985). Stereopsis and surface contours provide different constraints on 3D shape, and the strength of one cue might be expected to resolve the ambiguity of another. For example, stereopsis might serve to verify certain assumptions necessary to interpret monocular images, such as the angle of intersection of two contours. The results, reported in (Stevens & Brookes, 1987, 1988) and indicated by (2) in figure 1, were surprising: for the stimuli we used, which

involved planar surfaces, the stereo information had little influence on the end percept. For curved surfaces the story was quite different, for instance if the monocular information suggested a planar surface but the stereo suggested a Gaussian-shaped protrusion. We concluded that stereopsis provides strong constraints on the perception of surfaces *only where the second spatial derivatives of disparity are nonzero* (which correspond to regions of surface curvature and sharp discontinuities). Similar ideas have been put forward independently by at least two other groups of investigators (Gillam et al., 1984; Rogers, 1986). This work, initiated in the initial contract, led in the continuation of the contract to a basic reconsideration of the nature of depth from binocular disparity: depth is a reconstruction derived from second-derivative information.



(0) Initial ONR proposal
(1) Interconvertability of slant and depth
(2) Integration based on curvature and discontinuity features
(3) The reconstructive nature of stereo depth
(4) The nature of stereo features
(5) The construction of complex surfaces across cues

Figure 1. A graph of the major topics of research.

Following Werner's (1938) explanation of "depth induction" phenomena, we characterized the binocular depth reconstruction process as analogous to the reconstruction of brightness from luminance contrast information. In both cases the important information is carried by second derivative information, and in both cases the reconstruction is subject to a variety of artifacts. We explored the limits of this analogy in (Brookes & Stevens, 1989b) and indicated by (3) in figure 1, and made suggestions

regarding the origin of the differentiation steps, arguing in particular that one should not expect disparity processing to involve a circular-symmetric Laplacian-like operator.

We investigated the effects of the presence of surfaces on the perception of binocular depth and showed that the existence of surfaces can change the depth perceived from disparities. This work is reported in (Brookes & Stevens, 1989a) and indicated by (3) in figure 1. We have looked at how surface area affects resistance to noise and found that large areas are more resistant to noise. Also, we found that the features that appear to drive the depth percept do not appear to be the features which mediate detection of surfaces. This work is reported in (Brookes, 1988) and indicated by (4) in figure 1.

Finally, we extended our investigations to the integration of smoothly curved surface features defined independently by surface contours and binocular disparities, indicated by (5) in figure 1. In work reported in (Stevens, Lees and Brookes, in revision) we have demonstrated that the overall interpretation of 3D stimuli in which monocular and stereo cues conflict follows a more complex pattern than would be predicted by either winner-take-all or simple additive models of cue combination. For curved/planar combinations the integration appears to approximate a winner-take-all, or "cut and paste" model (see below). Thus, the monocular interpretation of a set of surface contours, whether planar or curved, tends to dominate the combined percept at locations in the display where the disparity pattern indicates planarity, while the binocular interpretation tends to dominate where the disparity pattern indicates curvature and where the monocular pattern indicates planarity. However, where both stereo and monocular interpretations indicate inconsistent surface curvature features, more complex resolution strategies are suggested, which may vary among different observers, and involve conscious attentive processing.

2. RESEARCH SUMMARY

The following is a summary of the results obtained by Stevens and Brookes from the present contract. We pursued the question of how 3D cues are combined in several different ways. We attempted to determine the set of primitives used in forming a depth percept, we compared the behavior of depth from disparities to that of brightness from luminance and finally we directly studied cases in which there was conflict between 3D cues.

2.1 The Depth Percept in Surfaces Depends on the Perceived Geometry of the Surface, not Directly on the Pattern of Disparities

The results described in Brookes and Stevens (1989a) show that binocular depth is computed subsequent to surface detection and that depth is computed from the surface descriptions. An experiment was performed to test this conjecture. The stimulus was a random dot stereogram with two different configurations. The first consisted of four slanted panels arranged roughly in a stairstep pattern. The slants of the panels were such that each panel had points of greater or lesser disparity than points on each other panel and yet had the overall impression of a set of slanted stairsteps. The other stimulus consisted of the same locations as the dots of the first stimulus but the disparities were randomized so that the disparity of each point was somewhere within the range of disparities of the first stimulus. The task, in the case of the paneled stimulus, consisted of showing one of the stimuli with a pair of probe points either on adjacent panels or on the outer pair of panels. For the random stimulus the same disparities were used which placed the probe points within the volume in depth. The subject was to decide which of the probe points was closer to the subject. The probe positions consisted of points that had equal disparities, points with greater disparities than those further up the stairsteps, and points with lesser disparities than further up the stairsteps.

The results of this experiment showed a significant difference between the depth judgments for surface versus random volume stimuli. For the random case (where the probes were embedded in a volume of stereo points) the relative depth of the probes were judged accurately in accordance with

their disparities. For the surface stimulus, however, the judgments for the probe points on the separated panels were consistent their being perceived as lying on a stairstep with little or no slant. This indicates an underestimation of the slants of the panels. For the adjacent panels, the depth of the probe points with larger disparity differences was judged correctly, but judgments for the probe points with smaller disparity differences and those with equal disparities again seemingly indicated underestimations in the slant of the panels.

If the depth of the pair of probe points were determined by a direct comparison of the disparities then the disparities of adjacent points should not affect the judgment. It appears that adjacent points which do not provide evidence of a surface do not affect the judgment. When the adjacent points are consistent with a surface, however, the judgment seems to be consistent with the properties of the perceived surface. This not only shows that the depth is reconstructed from surface discontinuities but also adds support to the conjecture that surface properties such as slant are inaccurately derived from disparities.

2.2 Depth is Analogous to Brightness in Effects Due to Reconstruction but not in Effects due to Spatial Lateral Inhibition

In the first funding period we established that depth is a reconstructed quantity for non-isolated binocular points. This reconstruction seems to be based on places in the image in which the second derivative is non-zero. These places, which include discontinuities and curvature features, were earlier found to be important in processing disparity information. Analogously, in the luminance domain, it has been established that there are mechanisms sensitive to discontinuities and extrema of luminance. Various contrast illusions in the luminance domain have counterparts in the disparity domain with similar behaviors. These facts suggested that depth might be processed in a manner similar to brightness. We found, however, that depth is analogous to brightness in effects due to reconstruction but not in effects due to spatial lateral inhibition.

Brookes and Stevens (1989b) explores this analogy by comparing known brightness illusions with their depth counterparts. Much work has been done with brightness, and the underlying mechanisms responsible for this processing are fairly well understood. Since only changes in luminance are detected, perceived brightness is largely a reconstructed quantity. The mechanisms involved in the detection of luminance differences induce lateral inhibition effects which take the form of illusory bands or spots at areas of changing contrast. If brightness and depth were completely analogous, depth would show some type of lateral inhibition effects as well as reconstruction effects.

Various types of illusions were compared to test specific parts of the analogy. Patterns were used that are directly analogous to patterns which exhibit brightness contrast effects in the luminance domain. Changes in luminance were mapped to changes in disparity. It was discovered that illusions due to reconstruction of brightness values have counterparts in depth perception but that those due to spatial lateral inhibition do not.

2.3 Surfaces are Detected on the Basis of Coherent Disparity Change Registered Prior to the Detection of Curvature and Discontinuity Features

The previous results brought up a more basic question: what constitutes a surface and how are surfaces detected? Related to this problem we found that surfaces are regions with an above threshold signal within a range of disparities and that surfaces are detected prior to the detection of salient surface features. Brookes and Stevens (in preparation) is concerned with problems in detecting and describing the surfaces that have been found to be so important. Two particular areas are addressed with further study suggested in certain areas. Both areas are concerned with how noise affects the detection of surfaces from stereopsis. Presumably some measure of the spatial coherence of the disparity field is used to determine that there is locally a surface that fits the disparity samples. In the absence of noise that measure reaches sufficiency with very few points: a very sparse collection of binocular points can be seen as lying on a smooth surface if their disparities vary correspondingly. With the addition of noise, it appears that sufficiency is reached by the density of coherent samples surpassing some critical level.

That is, with a certain density of points the surface should be perceived despite a substantial amount of noise (spatially uncorrelated disparity samples). This might be achieved by processes of facilitation and inhibition. With the combination of these processes the increase in strength of the surface is greater than linear. This suggests that a denser surface should have more resistance to noise than a sparse surface. The first experiment shows that this is the case. In this experiment, a random dot stereogram consisting of a planar surface parallel to the image plane is embedded in a certain percentage of points at random disparities. Subjects judged whether a surface was present in the image. The higher density surfaces were shown to be salient with a higher percentage of noise than the less dense surface

Another factor which affects the detectability of surfaces is the type of surface. That is, properties of the surface affect the detectability of the surface just as they affect the way depth is perceived from the surface. For example, surface edge information may be useful in detecting the presence of a surface. The ability to resist noise is a measure of the strength of particular surface being tested. The second experiment used this property to compare the salience of different surface types by comparing their resistance to noise.

2.4 Neither additivity models nor winner-take-all schemes account for cue integration phenomena

In Stevens, Lees and Brookes (in revision), we generated a series of stimuli in which different planar and curved patterns were independently defined by surface contours and by binocular disparity. The perceptual effects which we found resulting from the combination of these conflicting cues may be summarized as follows:

a) The monocular interpretation of a set of surface contours, whether planar or curved, tends to dominate the combined percept at locations in the display where the disparity pattern indicates planarity.

b) The binocular interpretation tends to dominate where the disparity pattern indicates curvature and where the monocular pattern indicates planarity.

c) Where both stereo and monocular interpretations indicate inconsistent surface curvature features, more complex resolution strategies are suggested, sometimes involving conscious attention to either the stereo or the mono interpretation, sometimes involving a compromise between both, but varying among observers and among presentations for the same observer.

d) Where both stereo and monocular interpretations indicate surface curvature features which are qualitatively consistent, but differ in amplitude, different observers show markedly different response patterns in a quantitative comparison task.

Binocular Depth from Surfaces vs. Volumes

Reprinted from the *Journal of Experimental Psychology: Human Perception and Performance* Vol. 15(3), 479-484.

Journal of Experimental Psychology:
Human Perception and Performance
1989, Vol. 15, No. 3, 479–484

0096-1523/89/$00.75

Binocular Depth From Surfaces Versus Volumes

Allen Brookes and Kent A. Stevens
Department of Computer Science
University of Oregon

Subjects were asked to compare the relative depths of two binocular targets embedded in different random dot stereogram backgrounds. The disparities of the background points were either randomized, corresponding to a scattering of points within a volume, or arranged according to a sawtooth (triangle-wave) disparity profile (i.e., a set of slanted planar surfaces separated by sharp depth discontinuities). When the targets were embedded in the random volume, their depths were perceived in accordance with their relative disparities. But when the target points were embedded in the sawtooth surfaces their depths were systematically misperceived in a manner predicted by the incorrect depth interpretation of the background points. Rather than seeing a sawtooth pattern, the background points resembled a staircase in depth, and the targets, which appeared embedded in different steps, were misjudged in depth accordingly. The effect suggests a distinction between the depth processing of isolated binocular features and those associated with continuous surfaces.

For distances measured radially from an observer, the *depth* associated with a given location is the difference in distance between that location and a given reference location. Depth, which is generally small compared to the overall reference distance, is often used to describe incremental distance variations, such as surface relief. *Apparent depth* is presumably the direct perceptual counterpart to this geometric quantity, so that the apparent three-dimensionality of viewed surfaces is usually expected to correspond to the determination of apparent depth for points across the given surface. There is, in principle, a direct geometric relationship between depth and binocular disparity, where the point of convergence of the two eyes provides a natural reference distance (see formulations in Foley, 1980; Mayhew, 1982). At least in the near field, apparent depth has been shown to be directly related to disparity and convergence (Foley, 1980; Morrison & Whiteside, 1984; Richards & Miller, 1969; Ritter, 1977, 1979). The visual system partially compensates for the dependency of depth on the square of the distance to the point of convergence (Ono & Comerford, 1977; Wallach, Gillam, & Cardillo, 1979). Foley has shown that systematic errors in binocular depth can be attributed to errors in the estimation of the apparent reference distance on the basis of an extraretinal convergence signal. Vertical disparities or eye movements have also been proposed as contributing to determining the geometric parameters of the binocular system necessary for recovering depth (Longuet-Higgins, 1982a, 1982b; Mayhew, 1982; Prazdny, 1983). It should be noted that whereas binocular disparity is often described in terms of absolute retinal positions, there is evidence that the effective binocular disparity is determined by differences between the two half-images, as suggested by our ability to maintain stable fusion despite retinal motion (Lappin, 1985; Steinman & Collewijn, 1980; Steinman, Levinson, Collewijn, & van der Steen, 1985).

It has been widely presumed that binocular depth across continuous surfaces is a straightforward extension of that associated with discrete binocular features. A continuous surface would present a rather dense sampling of binocular features, each contributing to the impression of depth at the corresponding surface location, probably on the basis of local disparity differences or contrast (Gogel, 1956, 1972; Gulick & Lawson, 1976). The importance of disparity contrast, and not absolute disparity, in determining apparent depth was first suggested by certain "depth contrast" effects (Pastore, 1964; Pastore & Terwilliger, 1966; Werner, 1938, 1942). A simple example of depth contrast is that of a central line at 0° disparity surrounded by flanking lines or dots that have disparities consistent with lying on a slanted plane: The central line will appear to slant away from the (apparently unslanted) frame. Depth contrast has been attributed primarily to the process of binocular fusion (e.g., cyclotorsion or shifts in effective correspondence; Nelson, 1977; Ogle, 1946), perhaps with the apparent frontoparallel, or zero-disparity, plane influenced by monocular cues (Harker, 1962).

Whereas disparity contrast seems necessary for the perception of apparent depth, recent observations suggest that it is not sufficient. Specifically, coplanar arrangements of binocular features, corresponding to slanted planes, have been found relatively ineffective in inducing apparent slant. Gillam, Flagg, and Finlay (1984) found that the slant of a plane is perceived much more rapidly when bounded by disparity discontinuities, and that, in their absence, depth develops with a slow time course similar to that reported in "aniseikonia" experiments (Ames, 1946). Mitchison and Westheimer (1984) also found that depth derives less effectively when the disparity features correspond to a coplanar arrangement (i.e., lying on a slanted plane). They found that the threshold for detection of apparent depth is elevated when adjacent binocular features are coplanar, and that the slant is particularly difficult to discern for certain arrangements, particularly those that mo-

This research was supported by Office of Naval Research Contract No. N00014-87-K-0321.

Correspondence concerning this article should be addressed to Kent A. Stevens, Department of Computer Science, University of Oregon, Eugene, Oregon 97403.

nocularly suggest an unslanted configuration, such as a square (McKee, 1983; Werner, 1937; Westheimer, 1979). The dominance of the monocular interpretation over constant disparity gradients was shown recently (Stevens & Brookes, 1988) for a variety of stereograms in which the distribution of binocular disparities corresponded to a slanted plane whose orientation was inconsistent with the monocular interpretation (e.g., as suggested by linear perspective). Given a sufficiently compelling monocular configuration, even very large contradictory disparity gradients are ineffective, provided they correspond to coplanar binocular features, and are presented in the absence of boundary disparity contrast.

The observations that binocular depth is dependent on the presence of disparity contrast and that depth is least reliably recovered from constant disparity gradients suggest an analogy between depth from disparity contrast and brightness from luminance contrast. Central to the analogy is the fact that binocular depth, like brightness, appears to be reconstructed across continuous regions bounded by contrast edges, as demonstrated by the depth analogue of the Craik-O'Brien-Cornsweet effect (Anstis, Howard, & Rogers, 1978). Other brightness analogues can be demonstrated; for example, a constant disparity gradient induces a complementary slant in a ring of constant disparity (Stevens, 1986; Stevens & Brookes, 1987)—an effect likely related to depth induction first observed by Werner (1938). See Brookes and Stevens (in press) for a discussion of the limits of this analogy.

Several explanations have been offered for the observed insensitivity to low spatial frequency variations in disparity, including spatial lateral inhibition (Anstis et al., 1978; Tyler, 1983) and local processes that align retinal images prior to binocular fusion (Anderson & Van Essen, 1987). But whereas some low spatial frequency depth information is seemingly lost at an early stage of binocular processing, the various depth contrast effects just mentioned show that at least some of that information is subsequently reconstructed. But other than that this information demonstrates the existence of binocular depth reconstruction, little more is known about it.

Mitchison and Westheimer (1984) characterize depth as being derived from differences in local disparity contrast; they observe, for example, that lines that have the same disparity difference between themselves and their neighbors appear to be at equal depths. This accounts for a variety of phenomena involving coplanar binocular arrangements that exhibit little apparent depth. However, their explanation seems to us more closely tied to the local detection of surface curvature or discontinuity features that are based on disparity rather than on the overall reconstruction of depth. More generally, our experience with similar stimuli has been that features embedded in continuous surfaces assume the apparent depth of the immediately underlying surface, which might consequently cause the features to appear to be at different depths, on the basis of, for example, monocular cues (Stevens & Brookes, 1988). We examined here whether this tendency also holds for purely binocular stimuli.

The approach was to use two types of random dot stereogram (RDS) stimuli. In the first type, the dots were given systematically varying binocular disparities that corresponded to a triangle-wave surface (i.e., a series of linear ramps separated by sharp disparity discontinuities; see Figure 1A). In the second type of RDS stimulus, which served as a control, the points were distributed randomly in disparity so that they appeared to lie scattered throughout a volume of space (Figure 1B). The stimuli were presented with no visible disparity contrast with the margins of the display. The only contrast was within the RDS—either among the dots of the volume stimuli or, in the case of the triangle-wave stimuli, across the vertical margins between adjacent slanted planes. Of particular importance to this experiment is the fact that the impression of overall depth from the triangle-wave disparity profile is incorrect. The stimuli do not appear as a series of slanted planes at a common mean distance from the observer; rather, their slant in depth is underestimated, so that the sharp disparity discontinuities between planes induce an erroneous overall increase in depth across the pattern. The RDS is seen in depth immediately as an arrangement of slightly slanted planes, whose apparent overall depth variation is intermediate between a triangle-wave and a staircase profile. The magnitude of the staircase effect is at least as large as that observed in the depth analogue to the Craik-O'Brien-Cornsweet effect (Anstis et al., 1978). Given these two stimuli, subjects were asked to compare the relative depth of two embedded target points that could be readily discerned from the other points of the RDS.

The intention of this experiment is to demonstrate a dependence of binocular depth on the presence of continuous surfaces. The stimuli are intended to be purely binocular, as afforded by random dot stereograms containing no monocular surface features. It is conceivable that the fused stereogram contains residual monocular depth or slant cues that might influence the results. For example, dot density was uniform across the stereograms, and the individual dots were all the same size (slightly less than 1′); both these facts indicated a stimulus equidistant from the observer, contrary to the binocular interpretation. These influences, if measureable, would apply equally to all RDS stimuli, and would presumably serve to reduce the impression of varying depth. The more important effect pursued here is the influence of surfaces on the apparent depth of embedded target points.

Method

Apparatus. The RDS stimuli were generated by a Symbolics 3675 Lisp Machine and displayed on a Wheatstone-style stereoscope consisting of a pair of optically flat front-surfaced mirrors and Tektronix 634 monochrome displays. The monitors were 94 cm from the observer, as measured along the optic axis from eye to screen, and were viewed with a convergence angle consistent with the observation distance. The stimulus stereogram subtended approximately 7° and consisted of luminous points against a dark background; the stereoscope was viewed in darkness.

Stimuli. The triangle-wave surface stimuli consisted of 2,000 points whose disparities corresponded to four slanted planes, each subtending 1.8° horizontally by 5.8° vertically. Disparity varied linearly across each slanted plane and discontinuously across the vertical margins between adjacent planes. The overall disparity range was −1.53′–6.13′, well within Panum's fusional limit. The disparity gradient across each plane corresponded to one of two slants in depth, varying either 4.6′ or 6.1′ over the 1.8° width of the plane (see disparity profile in Figure 2).

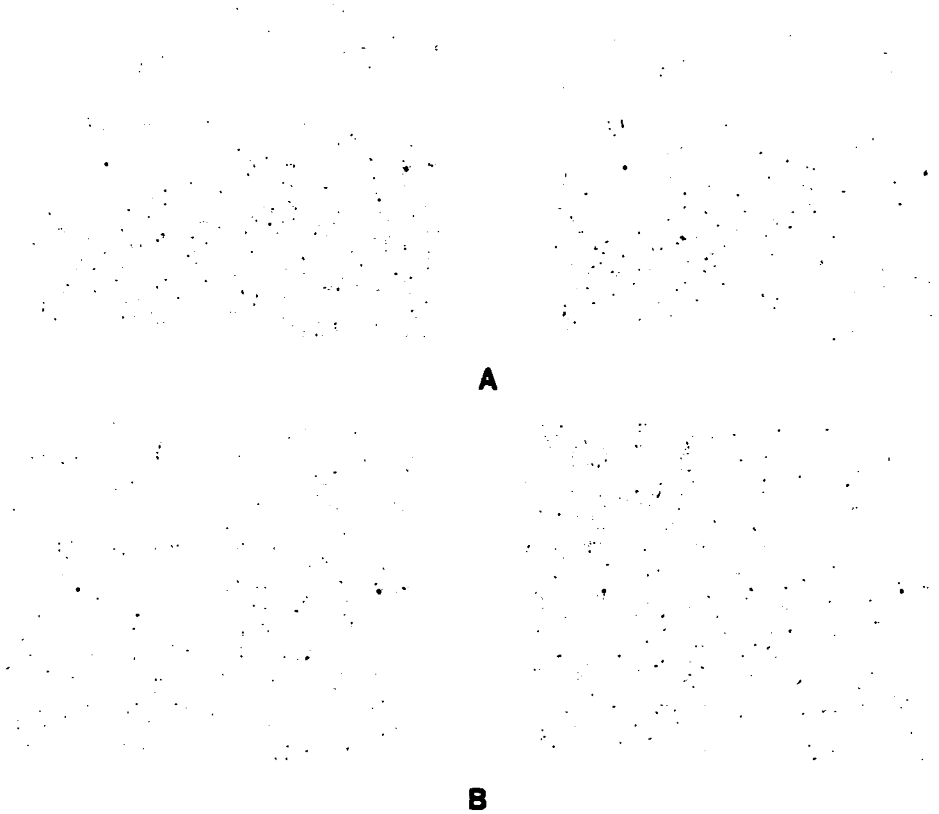


Figure 1. RDS stimuli similar to those used in the experiment. (In A the disparities correspond to a triangle-wave surface, but seen as having an overall staircase variation in depth. In B the disparities are distributed randomly, giving the appearance of a volume of points.)

Superimposed onto the RDS were two target points, each subtending 3′ so as to be distinguishable from the RDS points. The two target points had binocular disparities that matched the triangle-wave profile at its projected location, so that each target appeared to lie flush with the surrounding RDS surface. The two targets were positioned on the horizontal meridian to the left and right of the vertical meridian. The targets were embedded in either the central two planes (separated by 2° and one depth edge) or the outer two planes (separated by 6° and three intervening depth edges). We will refer to these as the near- and far-separation conditions. Figure 1 shows the two targets in the far-separation condition. For each of the two separations, the targets could appear in slightly different lateral positions on their corresponding slanted planes, so that four different relative disparities would result, specifically ±1.5′ and ±3.1′. Geometrically, a positive disparity difference corresponded to the condition in which the left target was nearer than the right, and a negative disparity difference corresponded

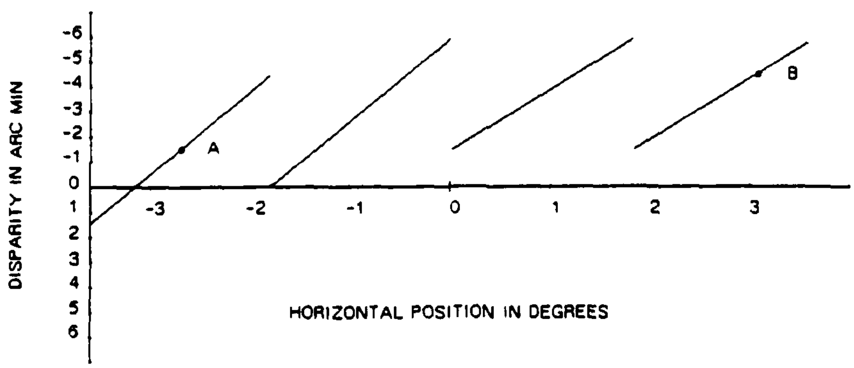


Figure 2. Disparity profile of the triangle-wave surface shown in Figure 1A. (A binocular target at location A tends to be seen as nearer than a target at location B, despite their relative disparities. Crossed disparities are negative in this figure.)

to the condition in which the right target was nearer than the left. The smaller disparity difference (±1.5′) was chosen empirically to be a challenging relative depth task for targets separated by 6° amid the other "distractor" points of the RDS. Altogether six combinations of binocular disparity were provided: the four combinations just described, and two conditions in which the two targets had equal disparity but different locations on their respective surfaces. Note that because the two targets appeared on ramps of differing disparity gradient, the relative depth judgment could not be deduced merely from their relative placement on the underlying planes. The mean disparities of the ramps were chosen in order to accommodate the range of relative target disparities.

A second experimental series was performed with the two targets embedded in random dot volume stimuli (Figure 1b). In this case, the same 2,000 dots were given random disparities within the same overall disparity range of −1.53–6.13′ used before. The dots that constituted the volume stimuli were fused readily, and they immediately appeared to define a volume of distinct points that were scattered in depth. The same six combinations of target location were used in these volume stimuli in conjunction with the two disparity senses (normal and reversed). Unlike the triangle-wave surface stimuli, in which the targets appeared to lie on surfaces in depth, the targets in this series appeared to float in space amid a random field of other three-dimensional points.

Procedure. Five experienced subjects participated in the experiment, 4 of whom were naive to the nature of the experiment; all had good stereo vision. In each trial the stimulus RDS was presented for 1,000 ms without the two targets, followed by an additional 750 ms, during which the target points were superimposed on the RDS. The subjects were told that they would see a pair of target points embedded in either a configuration of surfaces or a volume of points, and that they were to decide quickly but reliably which of the two targets appeared closer to the subject. The subject indicated the left or the right target by pressing the corresponding button on a mouse. The subjects were not given feedback about the accuracy of their judgments.

The experiment consisted of two series of trials: the triangle-wave surface stimuli (Figure 1A) followed by the random volume stimuli (Figure 1B). Each series consisted of 120 trials presented in random order: five repetitions of each of 12 distinct stimulus conditions, each presented for two choices of RDS disparity sense (normal and reversed, the latter of which served to reverse the direction in which depth increased in the apparent staircase). Note that the disparity reversal was for the entire stereogram, including the targets. The 12 conditions comprised six choices of position for the two targets and two target separations (near and far). Subjects were given learning trials without feedback until they indicated that they were comfortable with the task.

Results and Discussion

For the volume stimuli, the relative depth of the two targets was judged reliably for both the near (2°) and far (6°) separations. The *far-separation case*, not surprisingly, produced slightly more errors, particularly when the targets differed by only ±1.5′ in disparity (with 17% of the errors of the corresponding trials). In comparison, when the targets differed by ±3.1′ in disparity, their relative depth was judged accurately (with 1% of the errors of the corresponding trials) despite the large separation and the many intervening depth points.

The performance was quite different when the target points were embedded in the triangle-wave stimulus. An ANOVA was performed to test the main effects of (1) the surface versus volume background, (2) target separation, and (3) disparity difference. The presence of the surface was found to be significant, $F(1, 4) = 22.84$, $p < .05$. For the far-separation case, subjects had a strong tendency to make relative depth judgments consistent with the targets lying on separate planes in depth arranged as a staircase (rather than as a triangle-wave profile of slanted planes), despite the contradictory depth ordering implied by their disparities. For targets that were separated by only 2° (and lying on adjacent planes), the depth judgments were more in accordance with disparity, but were still judged contrary to disparity in 22% of the trials, in comparison to 5% for the volume stimuli. This corresponds to the subjective impression that the illusory staircase is relatively weak over adjacent step discontinuities and is most apparent when judging the relative depth of two points separated by several step edges. The cases most consistent with the illusory staircase involved targets separated by 6° (three intervening step discontinuities) and 1.5′ in disparity: The targets were seen in depth according to the apparent staircase and contrary to their disparity difference in 90% of the trials. Even for targets with a disparity difference of 3.1′, their relative depth was contrary to disparity in 66% of the trials, compared to 1% in the corresponding volume stimuli.

In Table 1, the data are collapsed across disparity reversals and presented in a manner that emphasizes the degree to which the responses were consistent with the staircase depth interpretation. As a basis for comparison, the bottom row shows how relative depth would be judged if based exclusively on binocular disparity. The data are presented with the convention that the apparent staircase increased in depth from left to right, so that a positive disparity difference would be consistent with the staircase. Note that the two conditions in which the targets had 0° disparity difference are presented together in the center column. For the volume stimuli there would be no expected bias (hence the .5 prediction), but for the surface stimuli we expected a bias if the targets were seen as lying at different depths on the apparent staircase. Note

Table 1
Fraction of Depth Responses as a Function of Disparity Difference of the Two Target Points, for Combinations of Target Separation and Surface Versus Volume

Variable	Disparity difference −3.10	−1.50	0.00:0.00	1.50	3.10
Predicted fraction	0.0	0.0	0.50:0.50	1.0	1.0
Volume stimuli					
Near-separation	0.02	0.08	0.36:0.56	0.96	0.98
Far-separation	0.02	0.14	0.40:0.42	0.80	1.0
Surface stimuli					
Near-separation	0.10	0.34	0.88:0.64	0.98	0.98
Far-separation	0.66	0.90	0.96:0.90	0.98	0.98

Note. The disparity differences are indicated in arc minutes. The central column shows the two conditions under which the targets had equal disparity. The top row shows the fraction of depth judgments predicted purely on the basis of their relative binocular disparities, hence 0.5 for the two cases of equal disparity. The numbers shown are the fraction of judgments consistent with the illusory staircase in depth seen in the sawtooth surface stimuli, where 1.0 would indicate all depth judgments corresponding with the targets lying on separate levels of the apparent staircase.

that the data for surface stimuli show an orderly bias towards the staircase interpretation, particularly for the far-separation case.

It has been demonstrated by many studies that depth can be derived readily from disparity contrast for spatially isolated targets. The targets in these stimuli were not isolated, however. It has been shown that binocular points in close proximity, separated by less than about 6–8′, exhibit depth averaging and spatial attraction and repulsion effects (Mitchison & McKee, 1985; Westheimer, 1986; Westheimer & Levi, 1987). In our experiment, the dot density was such that several background points could be expected to lie within approximately 6′ of each target. It is therefore conceivable that the relative depth of the targets was perturbed by adjacent RDS points, which contributed to the observed error rates. These perturbations, of course, would not have systematically influenced the target depths in the triangle-wave stimuli.

Probably more relevant is a second type of spatial interaction, which, as discussed earlier, tends to reduce apparent depth among coplanar binocular features. We expect that binocular depth is reconstructed across regions of continuous disparity change on the basis of the boundary conditions, such as the sharp disparity discontinuities between adjacent planes in the triangle-wave stimuli. Because we are relatively insensitive to the disparity gradient within the individual slanted planes, the reconstruction process erroenously accumulates an overall depth increase across subsequent planes, giving the impression of a staircase in depth. Because the targets were preceived as lying on the surfaces, their apparent depths were subject to errors of depth reconstruction.

Conclusions

The relative depth of a pair of isolated targets separated by a visual angle of several degrees and by several minutes of disparity can readily be determined from the targets' binocular disparities. In the present experiment, this was demonstrated by the volume condition, in which the two targets were embedded in a volume of distractor points. In the surface condition, the positions of the distractor points were held constant, but their disparities were distributed systematically, rather than randomly, in depth. Thus, the only difference between stimuli in the two conditions was in the disparity distribution of the distractor points. When the distractor points defined continuous surfaces, the relative depths of the embedded targets were no longer determined solely by their relative disparities. Rather, the target points acquired the depth of the embedding surfaces, and thus became subject to reconstructive errors in the perception of the surfaces.

The surface condition stimuli were designed to introduce an illusory impression of a staircase in depth, by capitalizing on the relatively greater perceived depth produced by sharp disparity edges than by continuous ramps of similar overall disparity contrast. This effect allowed us to demonstrate that depth judgments for target points on continuous surfaces are mediated by processes that access the reconstructed depth of the underlying surfaces, rather than being determined by their true disparity difference.

References

Ames, A., Jr. (1946). Binocular vision as affected by relations between uniocular stimulus-patterns in commonplace environments. *American Journal of Psychology, 59*, 333–357.

Anderson, C. H., & Van Essen, D. C. (1987). Shifter circuits: A computational strategy for dynamic aspects of visual processing. *Proceedings of the National Academy of Sciences, 84*, 6297–6301.

Anstis, S., Howard, I. P., & Rogers, B. (1978). A Craik-O'Brien-Cornsweet illusion for visual depth. *Vision Research, 18*, 213–217.

Brookes, A., & Stevens, K. A. (in press). The analogy between stereo depth and brightness contrast. *Perception.*

Foley, J. M. (1980). Binocular distance perception. *Psychological Review, 87*, 411–434.

Gillam, B. J., Flagg, T., & Finlay, D. (1984). Evidence for disparity change as the primary stimulus for stereoscopic processing. *Perception & Psychophysics, 36*, 559–564.

Gogel, W. C. (1956). Relative visual direction as a factor in relative distance perceptions. *Psychological Monographs, 418*, 1–19.

Gogel, W. C. (1972). Depth adjacency and cue effectiveness. *Journal of Experimental Psychology, 92*, 176–181.

Gulick, W. L., & Lawson, R. B. (1976). *Human stereopsis.* New York: Oxford University Press.

Harker, G. S. (1962). Apparent frontoparallel plane, stereoscopic correspondence, and induced cyclotorsion of the eyes. *Perceptual and Motor Skills, 14*, 75–87.

Lappin, J. A. (1985). A test of the role of retinal correspondence in stereo acuity. *Investigative Ophalmology and Visual Science: ARVO Abstract Supplement, 26*, 241.

Longuet-Higgins, H. C. (1982a). The role of the vertical dimension in stereoscopic vision. *Perception, 11*, 337–386.

Longuet-Higgins, H. C. (1982b). Appendix to Mayhew, J.: The interpretation of stereo-disparity information: The computation of surface orientation and depth. *Perception, 11*, 405–407.

Mayhew, J. (1982). The interpretation of stereo-disparity information: The computation of surface orientation and depth. *Perception, 11*, 387–403.

McKee, S. P. (1983). The spatial requirements for fine stereoacuity. *Vision Research, 23*, 191–198.

Mitchison, G. J., & McKee, S. P. (1985). Interpolation in stereoscopic matching. *Nature, 315*, 402–404.

Mitchison, G. J., & Westheimer, G. (1984). The perception of depth in simple figures. *Vision Research, 24*, 1063–1074.

Morrison, J. D., & Whiteside, T. C. D. (1984). Binocular cues in the perception of distance of a point source of light. *Perception, 13*, 555–566.

Nelson, J. (1977). The plasticity of correspondence: After-effects, illusions and horopter shifts in depth perception. *Journal of Theoretical Biology, 66*, 203–266.

Ogle, K. N. (1946). The binocular depth contrast phenomenon. *American Journal of Psychology, 59*, 111–126.

Ono, H., & Comerford, J. (1977). Stereoscopic depth constancy. In W. Epstein (Ed.), *Stability and constancy in visual perception: Mechanisms and processes* (pp. 91–128). New York: Wiley.

Pastore, N. (1964). Induction of a stereoscopic depth effect. *Science, 144*, 888.

Pastore, N., & Terwilliger, M. (1966). Induction of stereoscopic depth effects. *British Journal of Psychology, 57*, 201–202.

Prazdny, K. (1983). Stereoscopic matching, eye position, and absolute depth. *Perception, 12*, 151–160.

Richards, W., & Miller, J. F. (1969). Convergence as a cue to depth. *Perception & Psychophysics, 5*, 317–320.

Ritter, M. (1977). Effect of disparity and viewing distance on perceived depth. *Perception & Psychophysics, 22*, 400–407

Ritter, M. (1979). Perception of depth: Processing of simple positional disparity as a function of viewing distance. *Perception & Psychophysics, 25*, 209–214.

Steinman, R. M., & Collewijn, H. (1980). Binocular retinal image during active head rotation. *Vision Research, 20*, 415–429.

Steinman, R. M., Levinson, J. Z., Collewijn, H., & van der Steen, J. (1985). Vision in the presence of known natural retinal image motion. *Journal of the Optical Society of America A., 2*, 226–233.

Stevens, K. A. (1986, October). *Three-dimensional shape from two-dimensional contour.* Paper presented at the annual meeting of the Optical Society of America, Seattle, Washington.

Stevens, K. A., & Brookes, A. (1987). Depth reconstruction in stereopsis. *Proceedings of the First IEEE Conference on Computer Vision, 1*, 682–686.

Stevens, K. A., & Brookes, A. (1988). Integrating stereopsis with monocular interpretations of planar surfaces. *Vision Research, 28*, 371–386.

Tyler, C. W. (1983). Sensory processing of binocular disparity. In C. M. Shor & K. J. Ciuffreda (Eds.), *Vergence eye movements: Basic and clinical aspects.* Boston: Butterworths.

Wallach, H., Gillam, B., & Cardillo, L. (1979). Some consequences of stereoscopic depth constancy. *Perception & Psychophysics, 26*, 235–240.

Werner, H. (1937). Dynamical theory of depth perception. *Psychological Monographs, 49*, 1–127.

Werner, H. (1938). Binocular depth contrast and the conditions of the binocular field. *American Journal of Psychology, 51*, 489–497

Werner, H. (1942). Binocular vision, normal and abnormal. *Archives of Ophthalmology, New York, 28*, 834–844.

Westheimer, G. (1979). The spatial sense of the eye. *Investigative Ophthalmology and Visual Science, 18*, 893–912.

Westheimer, G. (1986). Panum's phenomenon and the confluence of signals from the two eyes in stereoscopy. *Proceedings of the Royal Society of London Series B. Biological Science, 228*, 289–305.

Westheimer, G. & Levi, D. M. (1987). Depth attraction and repulsion of disparate foveal stimuli. *Vision Research, 27*, 1361–1368.

Received April 1, 1988
Revision received July 22, 1988
Accepted September 9, 1988

The Analogy between Stereo Depth and Brightness

Reprinted from *Perception* Vol. 18(5), 601-614.

Perception, 1989, volume 18, number 5, pages 601 - 614

The analogy between stereo depth and brightness

Allen Brookes, Kent A Stevens ¶
Department of Computer Science, University of Oregon, Eugene, OR 97403, USA
Received 21 November 1988

Abstract. Apparent depth in stereograms exhibits various simultaneous-contrast and induction effects analogous to those reported in the luminance domain. This behavior suggests that stereo depth, like brightness, is reconstructed, ie recovered from higher-order spatial derivatives or differences of the original signal. The extent to which depth is analogous to brightness is examined. There are similarities in terms of contrast effects but dissimilarities in terms of the lateral inhibition effects traditionally attributed to underlying spatial-differentiation operators.

1 Introduction

Stereo disparity contrast can induce 'depth contrast' in a manner analogous to various well-known brightness contrast effects. A classic brightness contrast demonstration is shown in figure 1a, which shows a variant of Koffka's ring (Koffka 1935). A ring of uniform luminance is embedded in a background of constant luminance gradient. The variable contrast between the ring and its immediate background induces variable apparent brightness around the ring. Analogously, the stereogram in figure 1b consists of a ring of uniform disparity embedded in a background of constant disparity gradient. The ring appears slanted in depth in the direction opposite to that of the background gradient. Just as the brightness in figure 1a is dependent on luminance contrast more than on absolute luminance, so the apparent depth in figure 1b is dependent more on disparity contrast than on absolute disparity.

Depth contrast effects were first observed in simple stereograms in which a figure at zero disparity appears to slant in depth as a consequence of its surrounding context (Werner 1938, 1942; Pastore 1964; Pastore and Terwilliger 1966). Ogle (1946) suggested that during the fusion process, in the attempt to bring the context to zero disparity, cyclotorsion induces opposite disparity in the figure. Nelson (1977) later provided various experiments that ruled out cyclotorsion as the sole explanation, and furthered Werner's (1938) proposal that disparity *contrast* is responsible for the induction of apparent depth. In a manner analogous to the relationship between brightness and luminance contrast, the apparent depth in certain stereograms seems more reliably related to disparity contrast than to absolute disparities.

The analogy between depth and brightness has already been explicitly proposed in discussion of a stereoscopic counterpart of the Craik - O'Brien - Cornsweet illusion (Anstis et al 1978; Rogers and Graham 1983). In the luminance version of this illusion, two fields of equal luminance meet at a border whose profile is shaped like a double spur. The impression is of two homogeneous regions differing in brightness separated by a sharp step edge. In the depth version, one of the fields is seen as closer. The illusion demonstrates that depth information is extrapolated over extended regions bounded by sharp disparity edges, much like the extrapolation of brightness information away from intensity edges.

Brightness perception has been treated mathematically as the two-dimensional integration of a derivative-like retinal signal (Schiffman and Crovitz 1972; Arend 1973; Blake 1985; Arend and Goldstein 1987). If the luminance signal is conveyed to the

¶ Author to whom correspondence should be addressed.

cortex in terms of second derivatives computed by centre-surround operators in the retina (see below), any brightness illusions that result can be regarded as failures to achieve an accurate reconstruction of the incident signal, in part due to information lost by the initial derivative-like measurements (eg from thresholding).

Several brightness phenomena can be neatly described in terms of an empirically-measured spatial modulation transfer function (MTF) (Cornsweet 1970). The retinal receptive field presumed to be largely responsible for the overall shape of the MTF is traditionally modelled as a difference of Gaussians (DOG) (Rodieck and Stone 1965: Enroth-Cugell and Robson 1966). The resemblance of this circular-symmetric operator to the Laplacian of a Gaussian has been noted (Marr and Hildreth 1980), although the actual ratio of space constants (between excitatory and inhibitory Gaussians) in retinal DOGs is far too great to constitute a quantitative approximation to the Laplacian of a Gaussian (Robson 1983). Nonetheless, center-surround antagonism provides the qualitative effect of Laplacian filtering, and the component Gaussian receptive fields of the DOG achieves the effect of low-pass filtering, relative to the size of the operator. Lateral inhibition thus underlies both the insensitivity to low-spatial-frequency luminance variations and the relative sharpening of sensitivity to luminance discontinuities (both of which are demonstrated by the Craik-O'Brien-Cornsweet illusion). Lateral inhibition has also been invoked to explain other instances of diminished sensitivity to

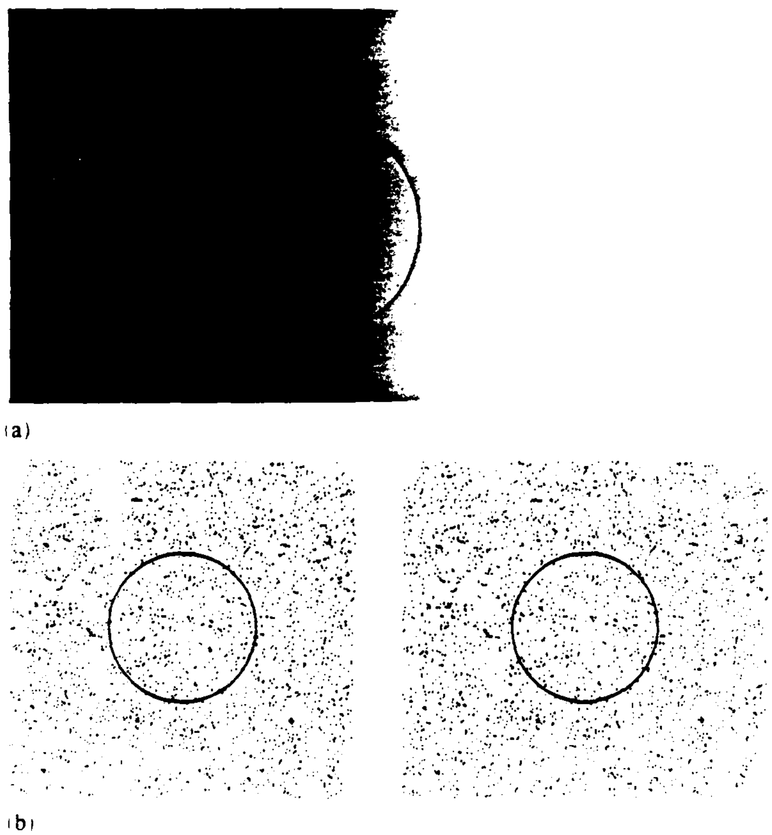


Figure 1. A variant of the Koffka ring. In (a) a ring of uniform luminance is embedded in a background of constant luminance gradient. In (b) the stereo disparity analogue presents a ring of uniform disparity against a background of constant disparity gradient. Note that the ring appears slanted.

low spatial frequencies, eg line spacing, line length, velocity, and motion in depth (MacKay 1973; Loomis and Nakayama 1973; Crovitz 1976; Regan et al 1986).

Stereo depth likewise exhibits an effective spatial MTF. Sensitivity to sinusoidal spatial modulations of stereo disparity is limited to a maximum of about 5 cycles deg^{-1}, with peak sensitivity at about 1 cycle deg^{-1}, and gradually diminishing sensitivity with decreasing spatial frequencies (Tyler 1973, 1975). The maximum sensitivity and high-frequency limits of our ability to see sinusoidal modulations in depth are consistent with independent evidence that continuous disparity distributions are spatially integrated within areas approximately 0.5 deg in diameter (Tyler and Julesz 1980). The gradual low-frequency falloff has been attributed to spatial lateral inhibition, eg by center-surround antagonism (Anstis et al 1978; Schumer and Ganz 1979; Tyler 1983; Schumer and Julesz 1984). It should be noted that two types of lateral inhibition can be expected in disparity processing: (i) spatial interactions, with summation or pooling of disparity signals within subfields and (center-surround) antagonism across spatially-separated subfields, and (ii) inhibition *across* disparity-tuned channels at a common location (Richards 1972; Tyler and Foley 1974; Nelson 1975; Marr and Poggio 1976; Julesz 1978; Westheimer 1986; cf Prazdny 1985). The high spatial-frequency limit would be evidence for spatial pooling or averaging of the disparity of closely-spaced features. Recently, Westheimer and Levi (1987) showed that, within about 4-6 min visual angle, binocular points show attraction in depth, and beyond that distance, repulsion in depth.

Do the substantial similarities between depth contrast and brightness contrast phenomena reflect similar processing strategies? We suggest that the observed similarities arise primarily from the fact that binocular depth and brightness are both reconstructed from (disparity or luminance) contrast, but that the analogy is limited because the corresponding contrast features are detected by fundamentally different strategies. The analogy is further limited by some evidence that the reconstruction strategies themselves also differ.

The discussion that follows gives instances where the analogy holds dramatically and obviously, and others where the analogy seems to fail. Where we report it fails, we are summarizing our experience over a variety of stimuli with several observers experienced in stereo observation. In the cases where the analogy holds, the effect in stereo depth is similar in strength to the traditional brightness effect. On the other hand, we have been unable to find a stereo counterpart for several other brightness effects. The breakdown of the analogy in these instances is regarded as significant in light of the strength and robustness of the original brightness effects.

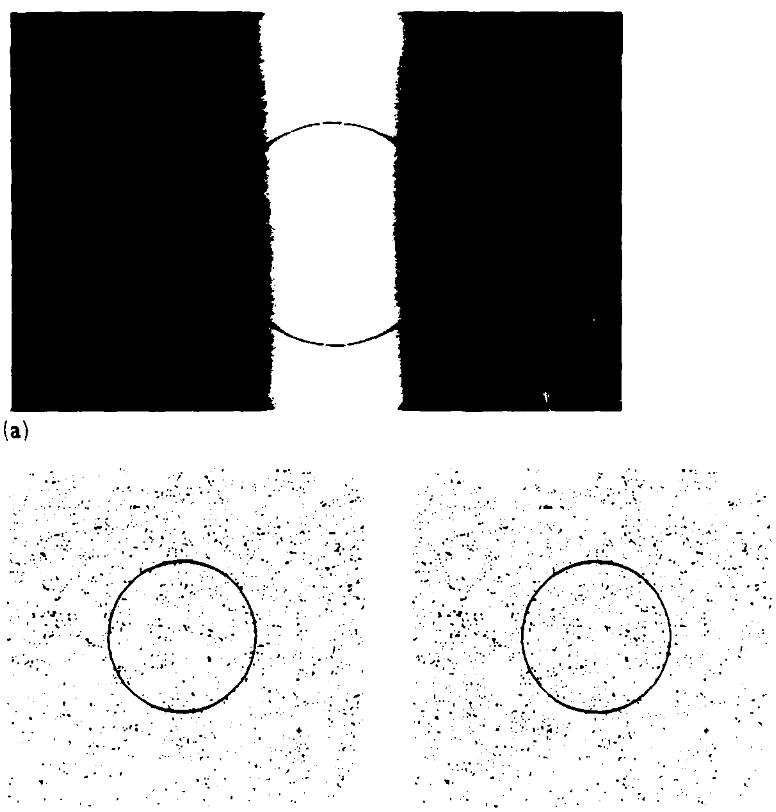
2 Brightness and depth effects associated with reconstruction

The Craik-O'Brien-Cornsweet illusion in stereo depth is compelling evidence that stereo depth derives from a process that reconstructs surfaces indirectly from boundary contrast. There are other demonstrations that depth derives from relative disparities, ie disparity differences within the binocular configuration, as opposed to absolute retinal disparities (Steinman and Collewijn 1980; Lappin 1985). The stereo analogue of the Craik-O'Brien-Cornsweet effect further shows that stereo depth is subject to errors in the integration of overall depth differences from subthreshold disparity variations. The difference in apparent distance from the observer to the left and right extremes of the pattern reflects a failure to incorporate the changes in very low spatial frequency into the accumulated depth variation over the pattern. Note that judging which side appears closer requires comparison of apparent radial distances. It is therefore remarkable that even with free eye movements observers cannot perform the task by comparing directly the disparities of the two regions. Clearly the distribution of

the surfaces in space is dominated by a (disparity) contrast-based reconstruction, seemingly in close analogy to the reconstruction of brightness in the original illusion.

The demonstration of simultaneous disparity contrast in figure 1 further shows that the perception of depth differences and of slant derives from local disparity contrast, eg across disparity discontinuities. Apparent slant across a continuous surface is no more reliably related to the local disparity gradient than is absolute depth to absolute disparity. The effect is thus closely analogous to brightness. One can readily generate further depth-induction counterparts to other brightness-induction demonstrations. For instance, just as two adjacent bars of the same luminance have different apparent brightnesses when presented against a luminance ramp background, adjacent lines of equal disparity appear at different depths when presented with a background of uniform disparity gradient (Mitchison and Westheimer 1984). These effects are not at all subtle: the ring in figure 1b appears dramatically slanted despite its uniform binocular disparity.

The local nature of the depth-induction effect can be demonstrated by means of a nonlinear background gradient, as shown in figure 2. In the luminance version (figure 2a) the constant-luminance ring is embedded in a Gaussian-shaped luminance profile. The brightness of the ring likewise varies with opposite sign to the background gradient. In the corresponding depth-version the constant-disparity ring is embedded in a Gaussian-shaped ridge in depth (figure 2b). The ring appears to curve in depth



(a)

(b)

Figure 2. Variant of the Koffka ring, similar to that in figure 1 but with a background with Gaussian profile. In a manner analogous to the variable brightness seen in the ring of uniform luminance in (a), the ring of uniform disparity in (b) appears curved in depth.

with induced curvature opposite to that of the background ridge. The curvature in depth induced in the constant-disparity ring is consistent with depth being dominated by the local disparity contrast, as is brightness by local luminance contrast. Note that in figures 1 and 2 the disparity gradient is horizontal to maximize the disparity contrast effect. Depth reconstruction effects in general have been shown to be anisotropic, stronger for horizontal compared to vertical gradients (Tyler 1973; Wallach and Bacon 1976; Rogers and Graham 1983).

Simultaneous *brightness* contrast is also seen when two squares of equal luminance are embedded in backgrounds of differing luminance. The square in the lighter background appears darker than the square in the darker background. Does it have a counterpart in stereo depth? The corresponding stereogram (figure 3) consists of two squares of equal binocular disparity embedded in regions of opposite disparity sign. For the analogy to hold, the square embedded in the negative-disparity background should appear farther away than that embedded in the positive-disparity background. But we find no corresponding depth difference in this configuration: the squares appear equidistant from the observer. The brightness contrast effect is often attributed to a logarithmic transformation of incident luminance (Cornsweet 1970): the compressive transformation results in differing effective contrasts prior to lightness reconstruction, and consequently differing apparent brightnesses. But no corresponding compressive transformation is found or expected for disparity because of the limited dynamic range of the disparity signal compared to that of the luminance signal (see Foley and Richards 1972; Foley 1980).

Another simultaneous-contrast effect is the apparent variation in brightness within a region of constant luminance induced by the contrast across its borders with adjacent regions. The familiar demonstration pattern consists of abutting rectangles of progressively higher luminance from left to right that produces a staircase luminance profile. Each rectangle appears distinctly lighter near the left margin and darker toward the right, an effect that is predicted by the spatial MTF (Cornsweet 1970). Figure 4 presents the analogous stereo stimulus: a staircase disparity profile. The apparent-depth profile is roughly analogous to the brightness version: the individual rectangles, despite their uniform disparity, appear slanted in depth. Although the depth increment across each sharp discontinuity is perceived rather accurately, apparent depth does not accumulate correctly over the staircase. As a result, the overall arrangement resembles a set of louvers, with the left side of each slanted rectangle appearing farther away than the right side.

The misperception of depth in the disparity staircase is predicted by the stereo MTF, much as the corresponding contrast sensitivity MTF predicts apparent brightness for

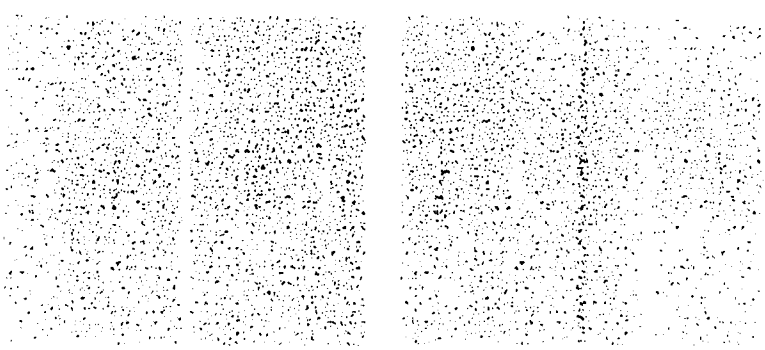


Figure 3. Stereo analogue for the brightness contrast effect. In this case there is no analogous effect.

the luminance staircase. But more is involved than is captured merely by a bandpass-filter model. A repeating triangle-wave disparity pattern, with constant mean disparity over the pattern, would be predicted on the basis of the MTF to be seen in depth veridically, but in fact is misperceived as a staircase depth profile (Brookes and Stevens 1989). Apparent depth increases across the pattern in a manner analogous to the accumulation of brightness reported for triangle-wave luminance profile sequences of Craik-O'Brien-Cornsweet edges (Arend et al 1971; Arend 1973; but see Coren 1983). Exceptions to the analogy concern the failure to observe perturbations to the apparent-depth profile in the vicinity of disparity discontinuities, the analogues of luminance effects traditionally attributed to lateral inhibition. We discuss this aspect of the analogy next.

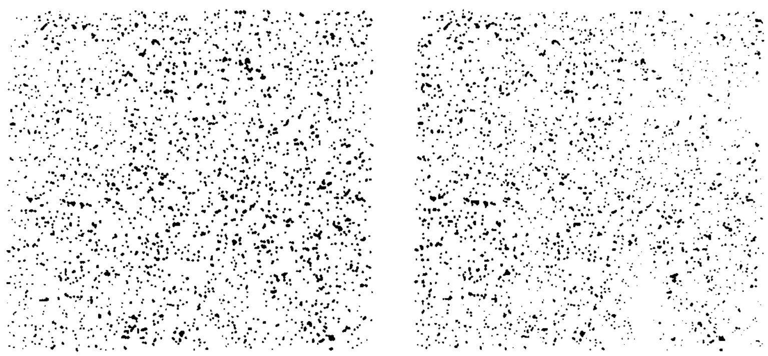


Figure 4. Stereo analogue for the simultaneous contrast effect. The staircase steps appear slanted but planar.

3 Effects associated with lateral inhibition

Several brightness phenomena appear directly to implicate neural mechanisms that might underlie aspects of the effective spatial MTF of the visual system. The first such mechanisms in the visual pathway are the retinal ganglion cells which, as mentioned, perform (spatiotemporal) derivative-like filtering by spatial lateral inhibition.

Mach bands are perhaps the most compelling illustration of lateral inhibition. The effect is an apparent creasing of the brightness profile where the corresponding luminance profile exhibits a sharp discontinuity in the second derivative. For example, dark and light lines are seen where a luminance ramp abuts the adjoining dark and light regions, respectively. Mach's proposal that the phenomenon derives from 'reciprocal action', ie lateral inhibition, of neighboring areas within the retina was later supported by direct neurophysiological recordings (Hartline and Ratliff 1957). Mach bands are robust over a wide range of luminance gradients, persist under focal scrutiny, and have measurable apparent width and amplitude, which can be related to the size of corresponding center-surround receptive fields in the retina (Ratliff 1965).

The Hermann grid illusion has been attributed to lateral inhibition, and specifically to center-surround receptive fields (Baumgartner 1960). The illusory spots seen at the grid intersections are consistent with the expected size of retinal center-surround receptive fields (Ratliff 1965; Spillman 1977). It should be noted that although the effect is likely due to lateral inhibition, it is doubtful that it arises solely from circular-symmetric retinal receptive fields; orientation-selective units have also been implicated (Levine et al 1980; Oehler and Spillman 1981; Wolfe 1984).

Several independent results would suggest that the features induced by lateral inhibition, if these are present, would be at least 6 min wide. Tyler (1973) showed that there is an upper limit of about 5 cycles deg^{-1} in the detection of sinusoidal variations

in depth in stereograms, which is equivalent to a half-cycle of 6 min. Mitchison and McKee (1985) have reported depth averaging for dots separated by less than about 6 min visual angle. Also, Westheimer and Levi (1987) have demonstrated a transition between attraction and repulsion in depth for targets separated by about 4-6 min. The attraction and repulsion effect is not particularly subtle: the magnitude of the apparent depth perturbation can be on the order of 1 min visual angle. Thus, if the spatial processes underlying these various lateral inhibition effects were to induce depth analogues to the corresponding binocular Hermann grid or Mach band stimuli, they should occur at approximately this scale and magnitude, or larger parafoveally.

In the depth version of the Hermann grid, consisting of a grid of squares above a background plane, the analogous effect would be illusory depth variations in the background at the grid intersections (either bumps or dips, depending on the disparity of the squares relative to the background grid). But the stereo analogue does not produce apparent illusory depth distortions at the grid intersections (Julesz 1965). Figure 5 shows a representative stereo depth version of the Hermann grid. The background surface appears uniformly planar, both where fixated and parafoveally.

Figure 6 shows the stereo analogue to the ramp-like luminance profile that generates the traditional Mach bands in brightness. The stereogram consists of a linear disparity gradient flanked by regions of uniform disparity. The depth analogue to a Mach band would be line-like ridges and troughs in depth where the disparity ramp abuts the regions of negative and positive disparity respectively.

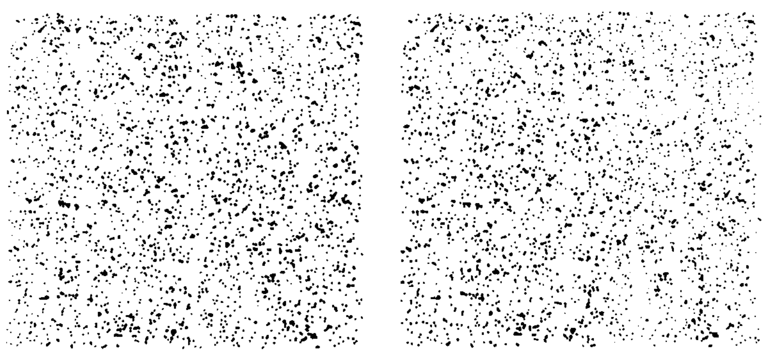


Figure 5. Stereo analogue of the Hermann grid. The background surface appears uniformly planar.

Figure 6. Ramp in depth between two unslanted planes. The corresponding luminance version induces Mach bands at the discontinuities where the gradient changes. In the stereo case there is no analogue to the Mach bands.

In examining for depth Mach bands, we used both dot and short-line stimuli, with densities similar to that in figure 6, and primarily varied the slope of the linear ramp region with disparity gradients that ranged from 1:8 to 1:3. For the moderately shallow 1:8 disparity gradient, the disparity varied over a total of 10 min visual angle across the length of the ramp. The spacing between adjacent dots or short lines was varied over a range of 2.3–6.1 min, with increments of about 0.8 min. Also, because of the known anisotropy between horizontal and vertical configurations (Tyler 1973; Wallach and Bacon 1976; Rogers and Graham 1983) both orientations were used for each spacing. No Mach-band-like depth effects were observed in stimuli where the ramp met the flanking level regions at a sharp crease, at any slope or orientation of stimulus. However, when the disparity profile was subtly modified to mimic Mach bands by the addition of slight ridges and troughs (0.8 min amplitude) at the margins between the ramp and the flanking regions, observers could readily discern the mock Mach bands.

A brightness effect similar to the Mach band is also to be found in a staircase luminance profile. In the immediate vicinity of each staircase step the brightness profile appears curved, an effect attributed to lateral inhibition (Ratliff 1965; Cornsweet 1970). The analogous depth effect would cause the uniform-disparity rectangles to appear *curved* as well as slanted in depth. But although the rectangles do appear slanted (figure 4), they appear distinctly planar. The disparity contrast across the step edge does not induce a local perturbation to the apparent surface in the vicinity of the edge.

Although subtle depth effects analogous to Mach bands and the Hermann grid effect might eventually be demonstrated, we find it noteworthy that they are not readily apparent, particularly given that discrete stereo features have been shown to exhibit substantial depth attraction and repulsion when brought into close proximity. This discrepancy suggests two possibilities, presuming the absence of the analogous effects is valid. Recall that Laplacian-like filtering enhances luminance changes and facilitates their subsequent localization, and that Laplacian-like filtering can be achieved by lateral inhibition or center–surround antagonism. One possibility, then, is that although some binocular mechanisms incorporate spatial lateral inhibition, those mechanisms are not involved in the detection of disparity change (ie depth edges). For example, lateral inhibition in the disparity domain is thought to be necessary for suppression of noise in stereo fusion and could cause depth contrast effects, but this is an interaction *among* disparity detectors, not necessarily a center–surround interaction (antagonism between excitatory center and inhibitory surround) within individual disparity detectors. Alternatively, lateral inhibition artifacts might be induced in depth by center–surround disparity-summating mechanisms but later suppressed at a subsequent stage of surface perception. We discuss these alternatives further below.

4 General discussion

The main points of the analogy between stereo depth and brightness contrast are (i) both brightness and depth appear to be reconstructions from boundary contrast features, and (ii) both luminance and disparity contrast features are seemingly defined by discontinuities or second spatial differences. The first point is supported by a range of contrast effects which establish the dependence of depth, like brightness, on the available boundary conditions, several of which were shown above. The second point is supported by many studies that demonstrate both the lack of direct correspondence between depth and disparity, and the relative insensitivity to constant disparity gradients. But the analogy has limits: while the reconstructions appear to embody similar computational principles, the detection of the underlying contrast or discontinuity events in the two domains is probably achieved by different methods. We first review the case regarding depth reconstruction.

4.1 *Depth reconstruction*

The notion that stereo depth is reconstructed indirectly from disparity contrast, much as is brightness from luminance contrast, is not particularly intuitive. The optical geometry of the two images has been shown by many theoretical analyses to support the direct pointwise computation of spatial information such as depth, slant, and absolute distance, provided that the necessary optical parameters are known from either retinal or extraretinal sources (Foley 1980; Mayhew and Longuet-Higgins 1982; Prazdny 1983). For simple binocular arrangements, often a pair of lines, the perceptions of depth, relative distance, and absolute distance are all rather accurately predicted by the direct geometric relationships, with systematic errors that can be attributed to misperception of the actual angle of convergence, differential magnification in the two eyes, and so forth (see review in Foley 1980). This evidence would suggest that a binocular observer is, at least for near objects, computing depth according to the optical geometry. Moreoever, apparent depth should vary approximately linearly with disparity and with the square of the observation distance (see Foley 1980 for a model for disparity targets at the fovea, Mayhew 1982 for a more general model that includes terms of eccentricity, and Cormack and Fox 1985 regarding stereograms). The influence of apparent viewing distance on apparent depth, an effect called 'depth constancy', is particularly apparent for small disparities and near observation distances (Ono and Comerford 1977; Ritter 1979; Wallach et al 1979).

It had been assumed, more or less tacitly, that such results would also apply to the points across a continuous binocular surface, eg with apparent depth varying according to the disparity at each surface point and apparent surface slant varying according to the disparity gradient (Mayhew 1982; Prazdny 1983).

Despite the elegance of the geometric equations and their predictions for simple binocular stimuli, other observations argue against a direct depth computation associated with each binocular feature, at least for those disparity distributions associated with continuous surfaces, whereupon the relative disparities become more salient than the absolute disparities within the configuration. As mentioned earlier, apparent depth remains invariant over differential retinal motions in the two eyes, which suggests that depth derives from the relative arrangement of disparities, and not from their absolute retinal coordinates (Steinman and Collewijn 1980; Lappin 1985; Regan et al 1986). Furthermore, the particular spatial arrangement of binocular features also matters, as demonstrated by depth attraction or repulsion between adjacent features and the diminished depth from coplanar arrangements of binocular features (McKee 1983; Mitchison and Westheimer 1984; Gillam et al 1984; Stevens and Brookes 1988). These observations together suggest an *indirect* relationship between disparity and depth for disparity distributions associated with continuous surfaces. In general, depth across continuous surfaces seems to derive *indirectly* from surface curvature features, which correspond to places where the second spatial differences of disparity are nonzero (Stevens and Brookes 1987, 1988), or in other words, where a gradient of relative disparities exists (Gillam et al 1988), which corresponds to differences of first differences (Mitchison and Westheimer 1984). Rogers (1986) has proposed that sensitivity to curvature in depth underlies the phenomenon of binocular depth constancy, again an indirect approach to surface perception from higher derivatives of the disparity field.

Thus the rather direct relationship between depth and disparity demonstrated for isolated three-dimensional features does not apply to the depth across continuous surfaces. In particular, when disparity varies linearly, as would occur in viewing a continuous slanted plane, apparent depth is determined by the disparity contrast across the borders of the plane relative to the background, if available. In the absence of border disparity contrast, the slant of the plane in depth is dominated by the monocular interpretation (Stevens and Brookes 1988).

In the luminance domain, brightness contrast effects reflect limitations in the ability of the visual system to reconstruct a luminance-related signal from measures of luminance change, presumably by interpolation (eg by lateral facilitation) within regions bounded by contrast features (Gerrits and Vendrik 1970; Davidson and Whiteside 1971; Arend 1973; Frisby 1979; Arend and Goldstein 1987). The stereo analogues suggest that binocular depth is likewise reconstructed, ie interpolated within regions bounded by disparity contrast features. Although the exact nature of the disparity features is not well understood, depth is elicited most effectively where the second spatial differences of disparity are nonzero, corresponding to surface discontinuity and curvature features (Stevens and Brookes 1987, 1988). And just as constant luminance gradients are effectively featureless and difficult to perceive, constant disparity gradients are similarly devoid of surface features and their interpretation in depth depends largely on the availability of disparity contrast, eg along their borders (Gillam et al 1984, 1988; Stevens and Brookes 1987, 1988).

4.2 *Discontinuity detection, spatial differentiation, and lateral inhibition*

The important binocular disparity features for surface reconstruction appear to correspond to loci where the second spatial differences of disparity are nonzero. Such features would be detected by measuring the second spatial derivatives of disparity. Spatial differentiation can be achieved effectively by center-surround lateral inhibition operators, a strategy that seems general to sensory processing. Whereas in the luminance domain the differentiation appears to be achieved by a circular-symmetric Laplacian-like filter, the known orientation anisotropy in sensitivity to disparity change would argue against a circular-symmetric operator for the corresponding detection of disparity features. Instead, one might postulate directional derivative operators composed of elongated receptive fields with lateral inhibition between adjacent subfields.

As discussed, there is evidence for the existence of very-short-range (several min visual angle) spatial lateral facilitation and inhibition in stereopsis. The effective spatial MTF of sensitivity to stereo depth also suggests lateral inhibition. But when one examines the stereo analogues of the traditional Mach band and Hermann grid, the expected lateral inhibition effects are not readily apparent. We see three alternative explanations.

First, the lateral inhibition effects in depth may simply have been more subtle than we allowed for in our explorations, or they were masked by the experimental design. But if the measured MTF for stereopsis is taken as an indication of the size of the underlying receptive fields, and if these receptive fields are presumed to summate disparities spatially in the conventional lateral-inhibitory manner, their effects would presumably not be particularly subtle.

The second alternative is suggested by the conventional wisdom that relative, if not absolute, binocular disparities are available after binocular fusion. Differentiation-like filtering of their spatial distribution would serve to detect possible surface features (discontinuities and other curvature events). As in luminance processing, the differentiation operator would produce patterns of activity that could be misinterpreted (eg Mach bands). But unlike luminance processing, which has only limited access to the original luminance signal, disparity processing could independently determine from the disparities in the immediate vicinity of each possible feature true features from artifacts. We see no way to test this alternative given the current state of understanding, or to distinguish it from the following alternative.

The third alternative is that disparity contrast features (edges and other curvature-related surface properties) are detected by processes that do not induce the characteristic lateral inhibition effects reported by others. Although both luminance

contrast and disparity contrast features seemingly require localizing changes in gradient (ie nonzero second spatial differences), they are unlikely detected by analogous operators. It would be disadvantageous to perform spatial differentiation by disparity-sensitive receptive fields which, by analogy, summate all disparity signals within small neighborhoods. To do so would be to blur not only in the two spatial dimensions of the image, but in depth as well. This would pose problems for the perception of transparent surfaces, where in a given visual direction at least two surface planes of disparities might be expected. It would be preferable to segregate perceptually those disparity signals that are likely associated with separate surfaces, prior to attempting to detect surface features. This alternative expects that those disparity distributions consistent with coherent surfaces (eg as measured in terms of local autocorrelation of disparity or local coplanarity) are treated differently than incoherent, or volume-filling, distributions (see evidence in Brookes and Stevens 1989).

We should note that an alternative method for computing a (directional) second difference is to perform two consecutive first-differences. The initial first-difference operation might be a consequence of compensating for uncontrolled disjunctive and conjunctive eye movements by shifting or remapping images (Anderson and van Essen 1987). As a result, positional information would be known only relatively (within each monocular image and between left and right images). The loss of absolute position information analogous to the loss of absolute luminance information causes simultaneous-contrast effects in motion perception as well as in stereopsis (Loomis and Nakayama 1973; Bowns and Braddick 1986; Rogers 1986).

If another first-difference operation were performed on the remapped images, the result would approximate a second directional derivative of the (motion or disparity) fields. Spatial differentiation might therefore be achieved by shifting rather than by convolution by lateral inhibition operators. There are, however, substantial control issues, such as determining the scale or locality over which a given shift is performed, and in spatially delimiting the application of a given shift.

Remapping or shifting is a particularly elegant solution to the problem of compensating for a spatially uniform error of unknown magnitude, where the relative signal is more reliable than the absolute. Anderson and van Essen (1987) expect the shifter to be controlled by a combination of feedforward (eg direct estimation of the local signal to nullify) and feedback (eg minimization of residual error or maximization of the measure of registration) strategies. Furthermore, if the magnitude and direction of the shift were determined locally for sufficiently small regions, the effect would remove or reduce constant gradients as well as spatially uniform terms. Local remapping would thus account for insensitivity to low-spatial-frequency disparity changes, as characteristic of differentiation operators. But it would also induce depth artifacts in the vicinity of disparity discontinuities characteristic of differentiation, which we did not observe. Moreover, the choice of control strategy for the shifter is particularly difficult for small populations of binocular features, such as used in Mitchison and Westheimer (1984, figure 5). Although remapping may contribute to the removal of low-spatial-frequency disparity information, it appears that the distribution of relative disparities is explicitly analyzed for planarity, as part of the extraction of surface discontinuity and curvature information.

In summary, stereo depth and brightness are analogous in that both are reconstructions: just as apparent brightness is dominated by the distribution of contrasts, stereo depth is dominated by the distribution of disparity contrasts. But the analogy does not extend to the corresponding (disparity and luminance) contrast-detection mechanisms. Depth contrast phenomena, like brightness contrast phenomena, stem from insensitivity to uniform gradients, as characterized by their respective spatial MTFs. In each case the visual system must reconstruct an approximation of the original distribution. The major

difference between the two domains seems to arise from the manner by which second spatial derivatives (or differences) are measured. Several observations argue against spatial differentiation of disparities in a manner analogous to the Laplacian-like operators applied to the luminance distribution. These include: (i) the horizontal-vertical anisotropy in depth reconstruction, (ii) the absence of analogous lateral inhibition effects (eg Mach band and Hermann grid phenomena), (iii) the plausibility that image registration or shifting processes (needed to control for dynamic positional errors between the two retinae) lose information about first differences, and finally, (iv) the implausibility of performing continuous differentiation on sparse, widely separated, discrete disparity features. Disregarding how the visual system measures second spatial derivatives of disparity (and to what extent our insensitivity to lower derivatives is a consequence of processes such as image registration), the *reconstruction* process, as far as we can tell, seems closely analogous in the stereo depth and brightness domains.

Acknowledgement. This research was supported by ONR contract N00014-87-K-0321.

References

Anderson C H, Essen D C van, 1987 "Shifter circuits: A computational strategy for dynamic aspects of visual processing" *Proceedings of the National Academy of Sciences of the United States of America* **84** 6297-6301

Anstis S, Howard I P, Rogers B, 1978 "A Craik-O'Brien-Cornsweet illusion for visual depth" *Vision Research* **18** 213-217

Arend L E, 1973 "Spatial differential and integral operations in human vision: Implications of stabilized retinal image fading" *Psychological Review* **80** 374-395

Arend L E, Buehler H N, Lockhead G R, 1971 "Difference information in brightness perception" *Perception & Psychophysics* **9** 367-370

Arend L E, Goldstein R E, 1987 "Lightness models, gradient illusions and curl" *Perception & Psychophysics* **42** 65-80

Baumgartner G, 1960 "Indirekte Grössenbestimmung der rezeptiven Felder der Retina beim Menschen mittels der Hermannschen Gittertäuschung" *Pflügers Archiv für die gesamte Psychologie des Meschen und der Tiere* **272** 21-22

Blake A, 1985 "Boundary conditions for lightness computation in Mondrian world" *Computer Vision, Graphics, and Image Processing* **14** 314-327

Bowns L, Braddick O J, 1986 "Evidence for a second derivative operator in motion parallax" *Investigative Ophthalmology and Visual Science, Supplement* **27** 181

Brookes A, Stevens K A, 1989 "Binocular depth from surfaces vs. volumes" *Journal of Experimental Psychology: Human Perception and Performance* **15** 479-484

Coren S, 1983 "When 'filling in' fails" *Behavioural and Brain Sciences* **6** 661-662

Cormack R, Fox R, 1985 "The computation of retinal disparity" *Perception & Psychophysics* **37** 176-178

Cornsweet T, 1970 *Visual Perception* (New York: Academic Press)

Crovitz H F, 1976 "Perceived length and the Craik-O'Brien-Cornsweet illusion" *Vision Research* **16** 435

Davidson M, Whiteside J A, 1971 "Human brightness perception near sharp contours" *Journal of the Optical Society of America* **61** 530-536

Enroth-Cugell C, Robson J G, 1966 "The contrast sensitivity of retinal ganglion cells of the cat" *Journal of Physiology (London)* **187** 517-552

Foley J M, 1980 "Binocular distance perception" *Psychological Review* **87** 411-434

Foley J M, Richards W, 1972 "Effects of voluntary eye movement and convergence on the binocular appreciation of depth" *Perception & Psychophysics* **11** 423-427

Frisby J P, 1979 *Seeing: Illusion, Brain and Mind* (Oxford: Oxford University Press)

Gerrits H H M, Vendrik A J H, 1970 "Simultaneous contrast, filling-in process, in information processing in man's visual system" *Experimental Brain Research* **11** 411-430

Gillam B J, Chambers D, Russo T, 1988 "Postfusional latency in stereoscopic slant perception and the primitives of stereopsis" *Journal of Experimental Psychology: Human Perception and Performance* **14** 163-175

Gillam B J, Flagg T, Finlay D, 1984 "Evidence for disparity change as the primary stimulus for stereoscopic processing" *Perception & Psychophysics* **36** 559-564

Hartline H K, Ratliff F, 1957 "Inhibitory interaction of receptor units in the eye of *Limulus*" *Journal of General Physiology* **40** 357-376

Julesz B, 1965 "Some neurophysiological problems of stereopsis" in *Proceedings of Symposium on Information Processing in Sight Sensory Systems* Ed. P W Nye (Pasadena, CA: Caltech) pp 135-142

Julesz B, 1978 "Global stereopsis: cooperative phenomena in stereoscopic depth perception" in *Handbook of Sensory Physiology* volume VIII, Eds R Held, H Leibowitz, H Tauber (New York: Springer) pp 215-256

Koffka K, 1935 *Principles of Gestalt Psychology* (New York: Harcourt, Brace and World)

Lappin J, 1985 "A test of the role of retinal correspondence in stereo acuity" *Investigative Ophthalmology and Visual Science, Supplement* **26** 241

Levine J, Spillman L, Wolf E, 1980 "Saturation enhancement in colored Hermann grids varying only in chroma" *Vision Research* **20** 307-313

Loomis J M, Nakayama K, 1973 "A velocity analogue of brightness contrast" *Perception* **2** 425-428

MacKay D M, 1973 "Lateral interactions between neural channels sensitive to texture density?" *Nature (London)* **245** 159-161

Marr D, Hildreth E, 1980 "Theory of edge detection" *Proceedings of the Royal Society of London, Series B* **207** 187-217

Marr D, Poggio T, 1976 "Cooperative computation of stereo disparity" *Science* **194** 283-287

Mayhew J E W, 1982 "The interpretation of stereo-disparity information: the computation of surface orientation and depth" *Perception* **11** 387-403

Mayhew J E W, Longuet-Higgins C, 1982 "A computational model of binocular depth perception" *Nature (London)* **297** 376-379

McKee S P, 1983 "The spatial requirements for fine stereoacuity" *Vision Research* **23** 191-198

Mitchison G J, McKee S P, 1985 "Interpolation in stereo matching" *Nature (London)* **315** 402-404

Mitchison G J, Westheimer G, 1984 "The perception of depth in simple figures" *Vision Research* **24** 1063-1074

Nelson J, 1975 "Globality and stereoscopic fusion in binocular vision" *Journal of Theoretical Biology* **49** 1-88

Nelson J, 1977 "The plasticity of correspondence: after-effects, illusions and horopter shifts in depth perception" *Journal of Theoretical Biology* **66** 203-266

Oehler R, Spillman L, 1981 "Illusory color changes in Hermann grids varying only in hue" *Vision Research* **21** 527-541

Ogle K N, 1946 "The binocular depth contrast phenomenon" *American Journal of Psychology* **59** 111-126

Ono H, Comerford J, 1977 "Stereoscopic depth constancy" in *Stability and Constancy in Visual Perception: Mechanisms and Processes* Ed. W Epstein (New York: John Wiley) pp 91-128

Pastore N, 1964 "Induction of a stereoscopic depth effect" *Science* **144** 888

Pastore N, Terwilliger M, 1966 "Induction of stereoscopic depth effects" *British Journal of Psychology* **57** 201-202

Prazdny K, 1983 "Stereoscopic matching, eye position, and absolute depth" *Perception* **12** 151-160

Prazdny K, 1985 "Detection of binocular disparities" *Biological Cybernetics* **52** 93-99

Ratliff F, 1965 *Mach Bands: Quantitative Studies on Neural Networks in the Retina* (San Francisco, CA: Holden-Day)

Regan D, Erkelens C J, Collewijn H, 1986 "Necessary conditions for the perception of motion in depth" *Investigative Ophthalmology and Visual Science* **27** 584-597

Richards W, 1972 "Disparity masking" *Vision Research* **12** 1113-1124

Ritter M, 1979 "Perception of depth: Processing of simple positional disparity as a function of viewing distance" *Perception & Psychophysics* **25** 209-214

Robson J G, 1983 "Frequency domain visual processing" in *Physical and Biological Processing of Images* Eds O J Braddick, A C Sleigh (Berlin: Springer) pp 73-87

Rodieck R W, Stone J, 1965 "Analysis of receptive fields of cat retinal ganglion cells" *Journal of Neurophysiology* **28** 833-849

Rogers B, 1986 "The perception of surface curvature from motion and parallax cues" *Investigative Ophthalmology and Visual Science, Supplement* **27** 181

Rogers B J, Graham M E, 1983 "Anisotropies in the perception of three-dimensional surfaces" *Science* **221** 1409-1411

Schiffman H, Crovitz H F, 1972 "A two-stage model of brightness" *Vision Research* **12** 2121-2131
Schumer R A, Ganz L, 1979 "Independent stereoscopic channels for different extents of spatial pooling" *Vision Research* **19** 1303-1314
Schumer R A, Julesz B, 1984 "Binocular disparity modulation sensitivity to disparities offset from the plane of fixation" *Vision Research* **24** 533-542
Spillman L, 1977 "Contrast and brightness illusions" in *Spatial Contrasts* Eds H Spekreijse, L H van der Tweel (Amsterdam: North-Holland) pp 45-49
Steinman R M, Collewijn H, 1980 "Binocular retinal image motion during active head rotation" *Vision Research* **20** 415-429
Stevens K A, Brookes A, 1987 "Depth reconstruction in stereopsis" in *Proceedings of the First IEEE International Conference on Computer Vision, London, June* (Washington, DC: IEEE Computer Society) pp 682-686
Stevens K A, Brookes A, 1988 "Integrating stereopsis with monocular interpretations of planar surfaces" *Vision Research* **28** 371-386
Tyler C W, 1973 "Stereoscopic vision: Cortical limitations and a disparity scaling effect" *Science* **181** 276-278
Tyler C W, 1975 "Spatial organization of binocular disparity sensitivity" *Vision Research* **15** 583-590
Tyler C W, 1983 "Sensory processing of binocular disparity" in *Vergence Eye Movements: Basic and Clinical Aspects* Eds C M Shor, K J Ciuffreda (Sevenoaks, Kent: Butterworths) pp 199-295
Tyler C W, Foley J M, 1974 "Stereomovement suppression for transient disparity changes" *Perception* **3** 287-296
Tyler C W, Julesz B, 1980 "On the depth of the cyclopean retina" *Experimental Brain Research* **40** 196-202
Wallach H, Bacon J, 1976 "Two forms of retinal disparity" *Perception & Psychophysics* **19** 375-382
Wallach H, Gillam B, Cardillo L, 1979 "Some consequences of stereoscopic depth constancy" *Perception & Psychophysics* **26** 235-240
Werner H, 1938 "Dynamical theory of depth perception" *Psychological Monographs* **49** 1-127
Werner H, 1942 "Binocular vision, normal and abnormal" *Archives of Ophthalmology, New York* **28** 834-844
Westheimer G, 1986 "Panum's phenomenon and the confluence of signals from the two eyes in stereoscopy" *Proceedings of the Royal Society of London, Series B* **228** 289-305
Westheimer G, Levi D M, 1987 "Depth attraction and repulsion of disparate foveal stimuli" *Vision Research* **27** 1361-1368
Wolfe J M, 1984 "Global factors in the Hermann grid illusion" *Perception* **13** 33-40

REFERENCES

Brookes, A. & Stevens, K.A. 1989a Binocular depth from surfaces vs. volumes. *Journal of Experimental Psychology: Human Perception and Performance*, **15(3)**, 479-484.

Brookes, A. & Stevens, K.A. 1989b The analogy between stereo depth and brightness contrast. *Perception*, **18(5)**, 601-614.

Brookes, A. & Stevens, K.A. The effect of surface features on detecting stereoscopic surfaces. In preparation.

Brookes, A. 1988 Extracting primitive surface descriptions with stereopsis. Ph.D. dissertation, Department of Computer Science, University of Oregon.

Gillam, B. Flagg, T., and Findlay, D. 1984 Evidence for disparity change as the primary stimulus for stereoscopic processing. *Perception and Psychophysics* **36**, 559-564.

Richards, W. 1985 Structure from stereo and motion. *Journal of the Optical Society of America* **2**, 343-349.

Rogers, B. 1986 The perception of surface curvature from motion and parallax cues *Investigative Ophthalmology and Visual Science ARVO Abstracts Supplement* **27**, 181.

Stevens, K.A. & Brookes, A. 1987 Probing depth in monocular images. *Biological Cybernetics* **56**, 355-366.

Stevens, K.A. & Brookes, A. 1988 Integrating stereopsis with monocular interpretations of planar surfaces. *Vision Research* **28**, 371-386.

Werner, H. 1938 Dynamical theory of depth perception. *Psychological Monographs* **49**, 1-127.

PUBLICATIONS, MEETINGS AND PERSONNEL

Refereed Publications:

Brookes, A. & Stevens, K.A. 1989 Local contour evidence of object occlusion *Proceedings of the 1989 International Conference on Image Processing,* Singapore, 823-826.

Brookes, A. & Stevens, K.A. 1989 Binocular depth from surfaces vs. volumes. *Journal of Experimental Psychology: Human Perception and Performance* **15(3)**, 479-484.

Brookes, A. & Stevens, K.A. 1989 The analogy between stereo depth and brightness. *Perception,* **18(5)**, 601-614.

Stevens, K.A. 1990 Constructing the perception of surfaces from multiple cues. *Mind and Language.* Special Issue (in press).

Stevens, K.A., Lees, M. & Brookes, A. Combining binocular and monocular curvature features. In revision.
1989

Stevens, K.A. & Brookes, A. 1990 A system for generation and modification of stereo disparity in images. *Proceedings of the 1990 International Conference on Automation, Robotics and Computer Vision,* Singapore.

Book Chapters:

Stevens, K.A. 1988 Visual object perception from a computational perspective. In *Visual object processing,* G. Humphreys & J. Riddoch, eds., 17-42. New York: Erlbaum.

Stevens, K.A. 1988 The perception of three-dimensionality across continuous surfaces. To appear as a chapter of a forthcoming book, S. Ellis, ed. New York: Lawrence Erlbaum Associates.

Technical reports:

Stevens, K.A. & Brookes, A. 1988 Binocular depth and the perception of visual surfaces. Technical Report 88-01, Office of Naval Research Contract Number N0014-87-K-0321

Stevens, K.A. & Brookes, A. 1988 Reconstruction of binocular depth across continuous surfaces. Technical Report 88-02, Office of Naval Research Contract Number N0014-87-K-0321

Meetings:

"Theory of depth reconstruction in stereopsis". Invited talk, University of California, Berkeley. May, 1987.

"The perception of three-dimensionality across continuous surfaces". Invited talk, Spatial Displays and Spatial Instruments. NASA-sponsored symposium and workshop. Asilomar, California. August, 1987.

"Integrating stereopsis and monocular depth". Thirteenth Annual Interdisciplinary Conference, Jackson, Wyoming. January, 1988.

"The reconstruction of continuous surfaces from stereo measurements and monocular inferences". Invited talk, Conference on vision and three dimensional representation, University of Minnesota. May, 1989.

"Surface shape, integration of 3D cues and computational problems". Keynote address, Fifth International Conference on Event Perception and Action, Miami University, Ohio. July, 1989.

"The role of curvature features in constructing perceived surfaces". Invited talk. Computational and Biological Models of Visual Processing. Trieste, February, 1990.

"The reconstruction of binocular depth". Invited talk. Reed College. April, 1990.

"On perceiving surfaces from monocular and binocular information". Invited talk. Department of Cognitive Science. University of California, Irvine. April, 1990.

Personnel:

Graduate students:

Allen Brookes
Marek Lees

PART II

INTEGRATING SPATIAL INFORMATION

Final Report ONR Grant N000-87-K-0321

Jacob Beck

This report summarizes research conducted between June 1, 1987 and September 15, 1987 and June 15, 1988 and June 30, 1990. I was on sabbatical leave between September 15, 1987 and June 15, 1988 and my research during this period was not supported by ONR.

CONTENTS

ABSTRACT

The research investigated the constraints or implicit assumptions employed by the visual system in the perception of tridimensional orientation in pictorial displays and how these constraints are applied, i.e., the algorithms used. We report studies on: (i) the effect of viewer distance on the perception of distance in pictorial displays, (ii) the constraints used by the visual system in perceiving a trapezoid as the perspective projection of a square, (iii) the constraints used in perceiving an obtuse angle, a parallelogram, and a sail figure as the orthographic projections of a right angle, a rectangle, and a sail, (iv) the algorithm used in perceiving a parallelogram as a rectangle, i.e., the computations applied by the visual system, and (v) the computations underlying the illusory perceptions of size occurring in orthographic projections. Two working hypotheses guided our research on the algorithms used. The first is that the system searches for the 3D orientation of a reference figure at which it matches a picture-plane variable. The search process is akin to what Perkins has called a direct computation (Perkins, 1983; Perkins & Cooper, 1980). It leads directly to the correct interpretation and does not involve either multiple paths or a search for interpretations that exhibit regularities. The second is that the computation is realized not by solving trigonometric equations but through internal representations of geometric operations. The computation is the geometric counterpart of a trigonometric calculation.

RESEARCH SUMMARY

1. Introduction

The mapping from three dimensions into two does not possess a unique inverse. The process of pictorial perception must therefore include rules for selecting one out of an infinite set of inverse transformations. How is the perceived 3D orientation of a surface in a pictorial display determined? Two general approaches have been proposed. The first is that the visual system selects an interpretation maximizing or minimizing a specific criterion, i.e., the Gestalt principle of Pragnanz (Koffka, 1935). Attneave (1972; Attneave & Frost, 1969) and Shepard (1981) suggest the visual system maximizes simplicity. An alternative view, the one we adopted, is that the visual system has developed inference rules which provide the necessary constraints. Examples of such inference rules are the interpretation of parallel curved contours as lines of curvature (Stevens, 1981; 1986), obtuse angles as right angles, (Perkins, 1972; 1973) and elliptic arcs as circular arcs (Barnard & Pentland, 1983).

The perceived 3D orientation of a surface has two degrees of freedom. Two constraints are therefore needed to recover perceived surface orientation from the projection of a surface onto a picture plane. We report studies on: (i) the effect of viewer distance on the perception of distance in pictorial displays, (ii) the constraints used by the visual system in perceiving a trapezoid as the perspective projection of a square, (iii) the constraints used in perceiving an obtuse angle, a parallelogram, and a sail figure as the orthographic projections of a right angle, a rectangle, and a sail, (iv) the algorithm used in perceiving a parallelogram as a rectangle, i.e., the computations applied by the visual system, and (v) the computations underlying the illusory perceptions of size occurring in orthographic projections.

2. The Analysis of Perspective and Orthographic Projections

The research we report was done within a larger theoretical view of how we perceive pictorial displays. This view is outlined in the following discussion.

We propose that the perception of 3D spatial orientation is the result of geometric transformations triggered by features of the pictorial pattern. The distortions arising from viewing a picture from an oblique direction need to be first corrected by processes akin to shape and size constancy (Pirenne, 1970; Perkins, 1973, Perkins & Cooper,1980; Wallach & Slaughter, 1986). The visual system is assumed to construct a 2D representation of the picture yielding the retinal image. When the picture plane orientation is correctly registered, the 2D representation constructed corresponds to the pictorial pattern. Features of the 2D representation are then interpreted to give tridimensional perceptions of orientation, shape, and size. Their interpretation is in terms of constraints or implicit assumptions employed by the visual system.

Two working hypotheses guided our research on the algorithms used. The first is that the system searches for the 3D orientation of a reference figure at which it matches a picture-plane variable. The search process is akin to what Perkins has called a direct computation (Perkins, 1983; Perkins & Cooper, 1980). It leads directly to the correct interpretation and does not involve either multiple paths or a search for interpretations that exhibit regularities. The second is that the computation is realized not by solving trigonometric equations but through internal representations

of geometric operations. The computation is the geometric counterpart of a trigonometric calculation.

3. Viewer Distance and the Perception of Distance in Pictorial Displays

There is a basic difference between the perception of objects in a real scene and in a pictorial scene. In a real scene, the visual system caries out to some approximation the inverse of the perspective projection of 3D objects onto the retina. Therefore, important factors in size and shape constancy are the distance of the viewer from an object and the slant of the object. The 3D perception of objects in a pictorial scene can not simply involve an inverse perspective transformation. The distance information necessary for carrying out such a transformation is not normally available. Since the units of distance in real space generally differ from the units of distance in pictorial space, real space and pictorial space are incommensurable. There is no difficulty in judging the distance between one's self and the picture in real units of distance and the distance between depicted objects in the picture in pictorial units of distance. However, it appears meaningless to judge the distance between one's self and an object in pictorial space.

Smith (1958) asked subjects to estimate the distance between two objects in a picture as well as the distance between a viewer an a point in the scene. He reported that the perceived interobject distances varied with the perceived distance of the viewer from a point in the scene. Smith, however, minimized the cues that one was looking a picture. In fact, he hypothesized that the size-distance relationship was found because of the highly realistic nature of the scene. We investigated how changing viewer distance affects perceived interobject distances in perspectively correct architectural drawings. The drawings were placed 24 and 72 inches from a viewer and it was apparent that one was looking at a picture. Viewers interobject distance judgments increased significantly only for 4 of the 24 drawings. For these 4 drawings, the mean perceived interobject distance at a viewing distance of 72 inches was 1.35 times that at 24 inches. Clearly, the size-distance computation does not affect the perceptions of distance in a pictorial scene and in a real scene in the same way.

4 Perceived Tridimensional Orientation of a Trapezoid

Since the process of perceiving pictorial representations fails to take into account the distance of the viewer from an object in pictorial space, the perception of pictorial space must have, at least in part, its own rules of interpretation. When the distance of the viewer is not taken into account, the slant of a square can be determined from its trapezoid projection if additional constraints are introduced. One type of constraint is to make assumptions about the position of the viewer. Given that a viewer's line of sight is normal to the center of the base of the trapezoid, we have shown that (1) a trapezoid can be the perspective projection of a square only if the height of the trapezoid is less than the width of the top, and (2) the slant of the square is given by the equation $\cos \sigma = h/t$ where σ is the slant of the square, h the height of the trapezoid, and t the width of the top of the trapezoid. These derivations hold whether the base of the perceived square is seen in the picture plane or behind the picture plane. The tilt (direction of slant) of the square follows from its symmetric convergence and is away from the observer along the line of sight.

It is important to point out that the visual system may at times utilize a constraint even when an assumption necessary for its derivation is violated. An example is the interpretation of a circle as a sphere. The projection of a sphere as a circle onto a planar surface occurs only when

the center of projection is perpendicular to the sphere, otherwise the projection is an ellipse (Pirenne, 1970). The visual system interprets a circle as a sphere even when the viewing angle is oblique. Perhaps this occurs because the projection of a sphere in a real scene is approximately circular even when the sphere is viewed obliquely. (The retina, unlike the picture plane, is a curved surface.) Whatever the reason, the visual system appears to interpret cues in accordance with established inference rules even when the viewing angle differs from the position in which the cue is mathematically valid.

An experiment examined two questions: (1) Would a subject judge a trapezoid to be a slanted square only when it is projectively possible? (2) How accurately can a subject judge the slant of a square from its trapezoid projection?

There were 18 stimuli. Nine stimuli were hard-copy images of computer generated projections of 9 squares slanted from 16.3 to 78.7 degrees floorwise. Nine stimuli were the same trapezoids except that their heights were made 10 percent greater than the top widths of the trapezoids. They could not be the projections of slanted squares viewed from a general position. The top row in Figure 1 shows the trapezoid projections of squares slanted floorwise 38.6 and 67.6 degrees when viewed from 15 in. and 34 in., respectively. The bottom row shows the corresponding trapezoids in which their heights were 10 percent greater than the top widths of the trapezoids. The trapezoids were placed upright on a stand positioned in one experiments at 2.5 times the correct observation distances. (The trapezoids were also presented at the correct observation distances but an experimental error made the data unusable. We plan to rerun this experiment.) A subject viewed the stimuli binocularly. Each subject's line of sight was normal to the center of the base of the trapezoid.

The 18 trapezoids were presented randomly 3 times to each of 10 subjects. A subject was instructed to try to see the trapezoids as surfaces slanted back in pictorial space. Each subject was then asked: Can this trapezoid be the projection of a square slanted away from you? If a subject said yes, the subject was asked to adjust a palm board to the perceived slant of the square.

When the trapezoids could be the projections of slanted squares, 61 percent of the judgments were that they were. When the trapezoids could not be the projections of slanted squares, only 19 percent of the judgments were that they were. The results support the hypothesis that subjects are sensitive to the height/top-width constraint on when a trapezoid can be the perspective projection of a square. Table 1 presents the means of subjects' slant judgments, their standard deviations, and the predicted slants for the 9 trapezoids which were perspective projections of slanted squares. Though there was considerable variability as indicated by the large

Table 1

Slant Judgments in Degrees

Stimuli	Mean Judgment	SD	Predicted Judgment
1	17.7	9.2	16.3
2	37.2	13.4	38.6

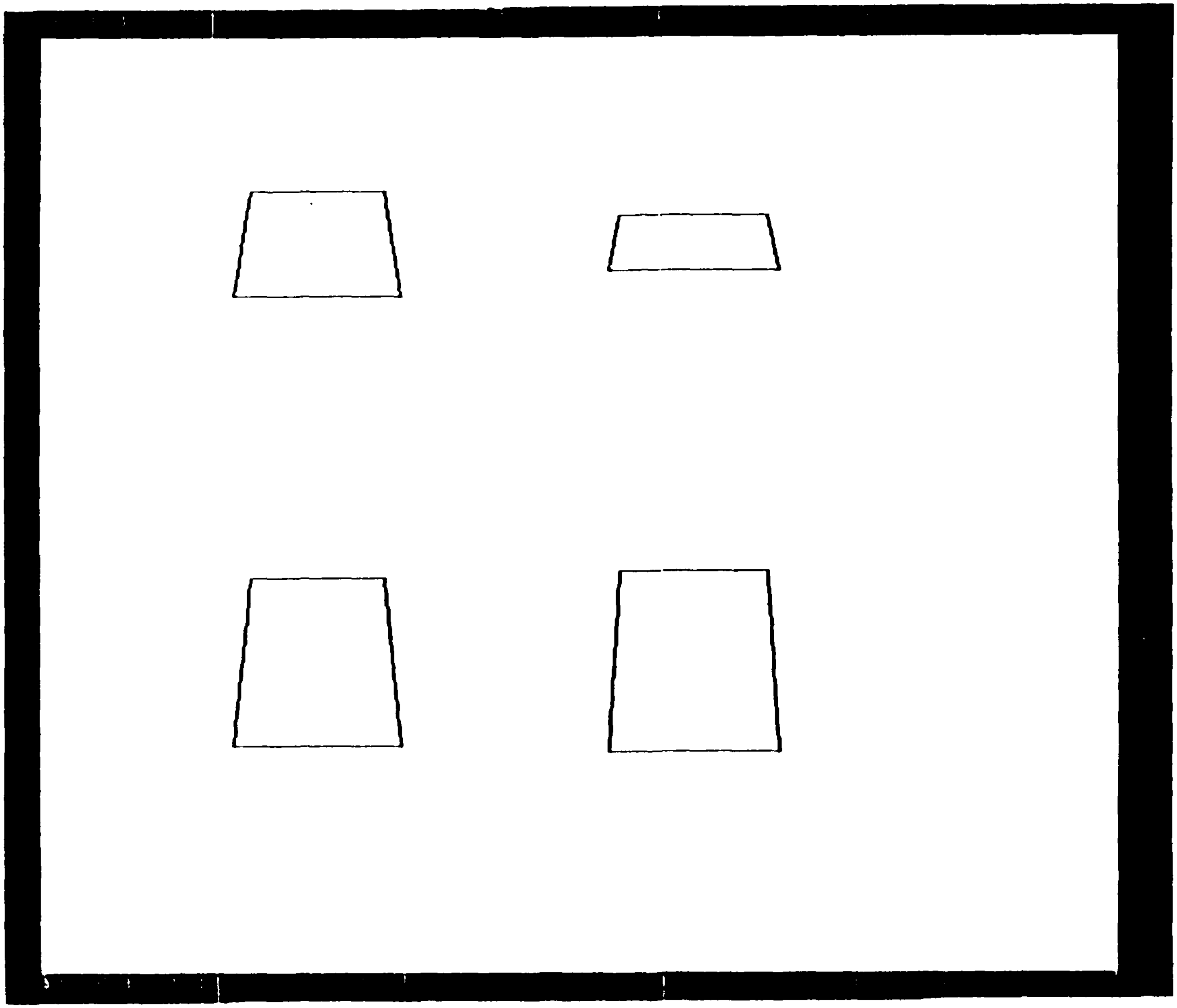


Figure 1

3	55.2	19.8	67.6
4	16.9	8.2	17.7
5	34.5	13.2	45.1
6	64.0	13.6	67.9
7	27.5	12.8	32.2
8	40.9	23.1	56.3
9	75.3	5.6	78.7

standard deviations, the mean slant judgments are remarkably accurate. This does not mean the visual system is solving algebraically a trigonometric equation. We believe the visual system solves the problem geometrically. What is suggested is that the visual system rotates in a mental analog of 3D space a square away from the frontal plane until the ratio of the height to the width of the top in the perspective projection is equal to that of the trapezoid stimulus (Shepard & Metzler, 1971). This calculation is independent of the distance of observation. Thus, subjects were able to make such accurate estimates although they were not at the correct distance of observation. We describe in Section 6.3 experiments to test that the algorithm used by the visual system involves geometric transformations.

5. Perceived Tridimensional Orientation of Orthographic Projections: Constraints

A strong perceptual tendency first pointed out by Mach (1959) is to perceive an obtuse picture angle as a right angle. Perkins (1972, 1973, 1983) has shown that the visual system imposes a right angle constraint when the constraint is projectively possible. As pointed out above, two constraints are necessary for fixing surface orientation, and even with the constraint that all angles should appear to be right angles, an additional constraint is necessary before surface orientation can be specified uniquely. The experiments investigated what additional constraints the visual system adopts in seeing an obtuse angle, a parallelogram, and the orthographic projection of a sail figure as slanted surfaces in pictorial space. The figures are readily seen as surfaces slanted in pictorial space. What is less evident is the second constraint adopted by the visual system. There are many possibilities.

5.1 Apparatus and procedure

The stimuli were displayed on a CRT monitor. The response apparatus described by Attneave & Frost (1969) was used. This apparatus allows a subject to adjust a luminous wand so that it appears normal to the perceived spatial orientation of the surface in pictorial space. When the base of the wand is centered on a stimulus surface, the subject feels as if he were objectively lining up the stick perpendicular to the surface. Slant and tilt of the wand can be independently adjusted and their values read from scales. In an orthographic projection it is always possible to reverse the near and far edges of a surface. Subjects, therefore, were asked to see the surface in a particular orientation. Subjects viewed the wand and the pictorial display binocularly. The slant and tilt were specified by the direction of the perceived normal to the surface. A surface in the frontal plane was perpendicular to the line of sight and at zero slant. Perceived tilt is the direction in which the surface was perceived slanted out of the frontal plane. Zero tilt corresponds to slanting the surface about the Y axis. The projection of the normal onto the frontal plane points at 3 o'clock.

5.2 Experiments

(i) Obtuse angle--Stevens (1981) proposed that the visual system slants a surface in the direction of the bisector of the range of permissible tilts. Stevens (1983) found that the relative line lengths may affect the perceived tilt of a surface suggested by intersecting lines. The surface was perceived tilted so as to equate the lengths of the lines. An experiment tested whether the perceived tilt of a surface suggested by an obtuse angle is affected by the relative lengths of the lines composing the angle. The experiments varied the size of the obtuse angle, the relative lengths of the lines composing the angle, and the orientation of the obtuse angle in the picture plane.

There were 18 obtuse angles in the experiment: six 110 degrees, six 125 degrees, four 145 degrees, and two 155 degrees. Stimuli with the same angles differed in their orientation in the plane. Subjects were instructed to see the obtuse angles as the edges of a surface oriented in 3-space. They were instructed to position the wand until it appeared normal to the surface defined by the angles. The angles were presented in a random order and each subject judged each of the angles three times. There were two parts to the experiment. In the first part, the lengths of the lines composing an angle were of equal length. In the second part of the experiment, the lengths of lines composing an angle were in a 3:2 ratio. The second part was run a week or more after the first part. Five subject served in the experiment.

The results are shown in Tables 2 and 3. The predicted slant and tilt judgments in Table 2 assume that the direction of perceived slant is in the direction of the angle bisector. The close agreement between subjects tilt judgments and the predicted tilt judgments indicate that the perceived direction of slant was in the direction of the angle bisector. The predicted slant and tilt judgments in Table 3 are based on the assumption that the direction of tilt is such to equalize the lengths of the lines composing the angle in 3-space. The asterisks are for cases in which if the obtuse angle is seen as a right angle there is no direction of slant which will equalize the line lengths.

Table 2

Slant and Tilt Judgments in Degrees

Equal Line Lengths

Stimuli	Mean Slant Judgment	Mean Tilt Judgment	Predicted Slant Judgment	Predicted Tilt Judgment
1	32	149	46	160
2	37	146	46	148
3	39	135	46	139
4	31	89	46	90
5	33	72	46	79
6	35	66	46	69
7	52	140	59	145
8	52	130	59	136
9	53	125	59	129
10	56	88	59	91

11	49	79	59	82
12	51	70	59	74
13	64	126	69	130
14	63	118	69	123
15	67	89	69	90
16	66	83	69	84
17	72	109	77	115
18	74	88	77	91

Table 3

Slant and Tilt Judgments in Degrees

Unequal Line Lengths (3:2)

Stimuli	Mean Slant Judgment	Mean Tilt Judgment	Predicted Slant Judgment	Predicted Tilt Judgment
1	29	159	51	142
2	35	144	51	130
3	38	142	51	121
4	41	88	51	72
5	33	84	51	61
6	34	70	51	51
7	52	142	69	124
8	53	134	69	115
9	53	130	69	108
10	53	92	69	70
11	53	84	69	61
12	54	76	69	53
13	62	130	*	*
14	61	124	*	*
15	67	89	*	*
16	69	85	*	*
17	71	113	*	*
18	73	90	*	*

The slant and tilt judgments in Table 3 are similar to those in Table 2. The results indicate that unlike intersecting lines, the perceived tilt of a 3D surface suggested by an obtuse angle is in the direction of the angle bisector for angles with lines of equal length and for angles with lines in a 3:2 ratio.

*(ii) **Parallelogram***--The orthographic projection of a slanted rectangle is a parallelogram. One possible constraint is that the perceived surface is slanted in the direction of lines that can be seen as normals to the surface. This presumption is of particular interest since we use it in testing our hypotheses about the algorithms employed by the visual system (see Section 6.3). Two experiments tested the hypothesis that the lines at the corners of a parallelogram are seen as

normals to the perceived 3D orientation of the surface. Two different parallelograms were used in the experiments. Five stimuli were used in the first experiment. The lines differed in their orientation and whether they pointed up or down. The 2D orientations of the lines were (3 o'clock being 0 and proceeding in a counterclockwise direction): 105• up, 120• up, 90• up, 90• down, and 60• down. Four stimuli were used in the second experiment: The 2D orientations of the lines were: 90• up, 90• down, 60• up, and 60• down. Eight subjects served in the first experiment and six in the second experiment. Each of the stimuli were presented five times.

Table 4 presents the results. The means of subjects' tilt judgments in both experiments were within 3 degrees of the 2D orientations of the lines in the corners of the parallelogram. This means that the perceived 3D direction of slant was around an axis of rotation in the plane that is perpendicular to the lines taken to be the surface normals. The slant judgments were less accurate. The stimuli with lines at 90 degrees have a greater slant than expected. This may be due to the vertical lines tending to pull the wand away from the subject's line of sight and toward the frontal plane.

Table 4

Slant and Tilt Judgments in Degrees

Stimuli	Lines	Mean Slant Judgment	Mean Tilt Judgment	Predicted Slant Judgment	Predicted Tilt Judgment
1	105• up	62	108	58	105
2	120• up	63	120	72	120
3	90• up	69	91	58	90
4	90• down	68	91	58	90
5	60• down	61	61	65	60
6	90• up	67	93	58	90
7	90• down	61	91	58	90
8	60• up	63	62	65	60
9	60• down	59	61	65	60

(iii) Sail figure--Figure 2 shows the sail figure with and without rulings, i.e., the straight lines connecting the curved contours. The parallel contours in the figure are interpreted as lines of curvature (Stevens (1981; 1986). Six subjects adjusted the wand to the perceived normal at three different points of the figure. The sail figures were presented with and without rulings. The wand coincided with the second ruling (first interior line from the top, stimuli 1 and 2), fourth ruling (stimuli 3 and 4) and sixth ruling (stimuli 5 and 6) when rulings were present. In previous experiments subjects were allowed full control over both slant and tilt. In this experiment, the subject could control only the slant of the wand toward or away from the frontal plane at a fixed tilt. The tilt of the wand was fixed at the angle bisector of the obtuse angle formed by a ruling and a contour at which the wand appeared pivoted.

Table 5 presents the results. Rulings were present for stimuli 1, 3, and 5, and absent for stimuli 2, 4, and 6. The predicted slant is given in the third column and is based on the assumption that the visual system interprets the obtuse angle formed by a ruling and contour as a right angle

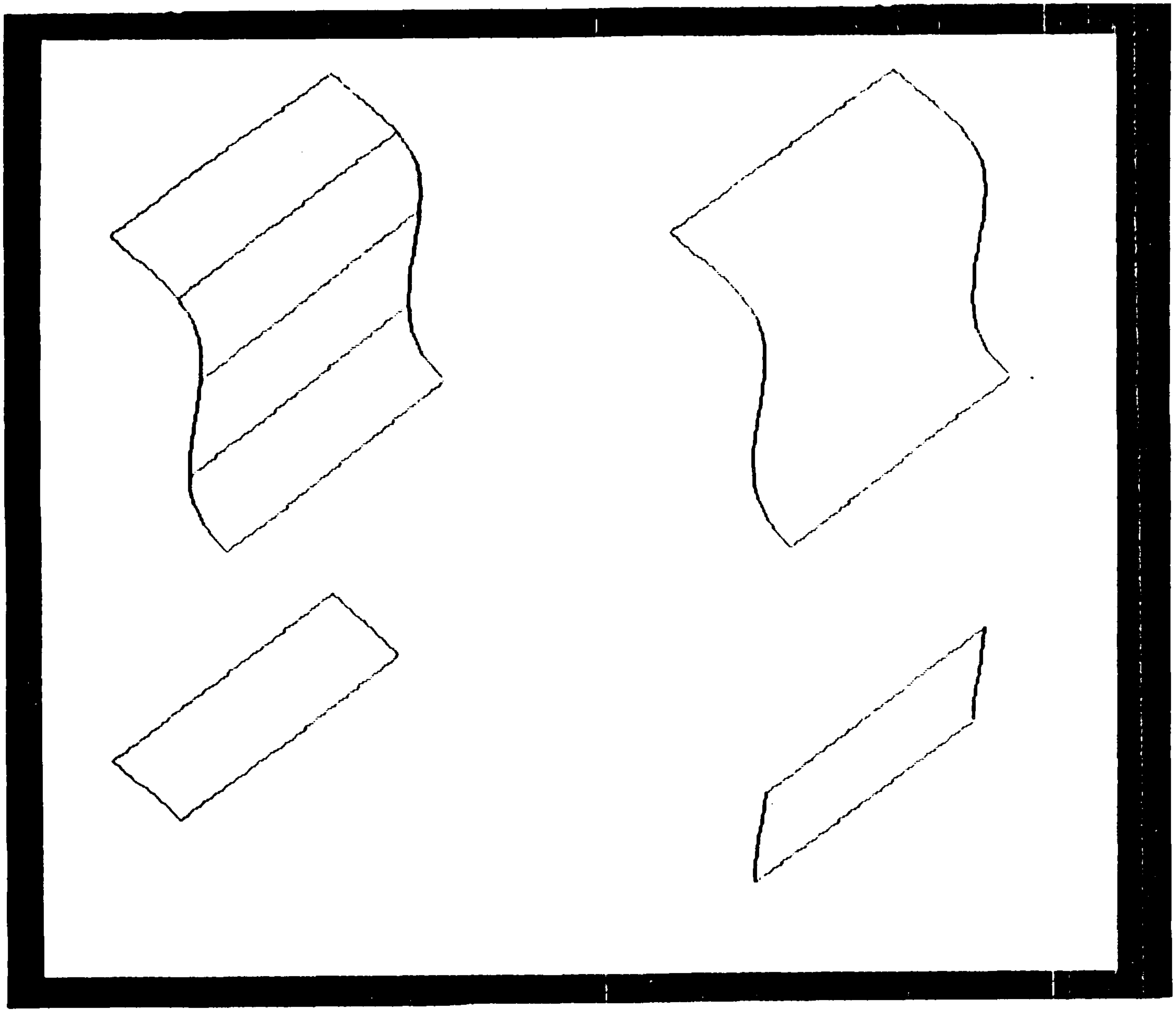


Figure 2

that is slanted in the direction of the angle bisector. The judged slants of the sail with and without rulings were similar. The results suggest that the visual system uses "virtual rulings" to establish a correspondence between the contours of the sail figure. The results also suggest that the slant of the sail is obtained by approximating the sail figure with parallelograms. The bottom figures in Figure 3 show the top and third from the top sections composing the sail surface. The perceived 3D spatial orientations of these individual sections appear similar to their corresponding sections in the sail figure.

Table 5

Slant Judgments in Degrees

Stimuli	Mean Slant Judgment	S.D.	Predicted Slant Judgment
1	33	8.4	45
2	31	8.5	45
3	46	9.1	65
4	43	9.8	65
5	27	11.8	42
6	26	16.4	42

6. Orthographic Projections: Algorithms

Trigonometric equations for deriving surface orientation from certain constraints that might be adopted by the visual system can be found in Attneave & Frost (1969), and Stevens (1981, 1983). Our hypothesis is that the visual system solves the problem geometrically instead of algebraically. We hypothesized that the perceived spatial orientation of a figure in pictorial space is the consequence of a sequence of geometric transformations. The computational algorithm involves five stages:

(1) The visual system selects a reference figure based on an interpretation of the picture-plane figure.

(2) The reference figure is rotated in the picture plane until there is a correspondence between a feature of the reference figure and a feature of the picture.

(3) The visual system fixes an axis about which the reference figure is rotated. This fixes one degree of freedom.

(4) The reference figure is then rotated about the axis of rotation until a feature in the reference figure is equal to a feature in the picture. This fixes the second degree of freedom.

(5) If the lines of the projection of the reference figure onto the picture plane are not in correspondence with the lines of the pictured surface, the reference figure is rotated about the normal to its surface until the lines of the projected reference figure match the lines of the pictured surface. This fixes the orientation of the projected surface in

the picture plane.

We leave open the question whether the above processes are to be identified with mental rotation in Shepard's sense (Shepard, 1981).

6.1 Axis of rotation: applications of the hypothesis

(i) ***Parallelogram***--The visual system is assumed to select a rectangle or square as the reference figure. What is the axis of rotation that allows a parallelogram to be seen as a square or rectangle slanted in 3D space? The experiments reported in Section 5 show that the direction of slant may be fixed by lines that are seen as normals to the surface. This is illustrated in Figure 3. The top left, top right, and bottom left parallelograms in Figure 3 are identical. Vertical lines have been added to the corners of the top right parallelogram and 50 degree lines to the corners of the bottom left parallelogram. There is a strong presumption to see the lines as the projections of normals to the surface in 3D space. This fixes the direction of slant and resolves the projection ambiguity by providing the necessary second constraint. The direction of slant of the reference rectangle must be around an axis in the plane that is perpendicular to the line that is taken to be the projection of the surface normal. The reference rectangle is slanted until the orthogonal projection of the right angle in the rectangle approximately matches the obtuse picture angle of the projected surface. The reference rectangle is then rotated about the normal to its surface until the lines of the projected right angle match the lines of the pictured surface.

6.2 Size illusion

Converging lines in the perspective projection are associated with distance and signal the visual system to correct the diminishing retinal image size of distant objects. Figure 4 shows the perspective projection of a sinusoidal cylindrical surface which we refer to as a bench. The near and far probes are of equal size but subjects' judged the far probe to be larger. It is an over-simplification to make the illusion depend solely on perspective cues. A similar illusion occurs with an orthographic projection. Figure 5 shows the orthographic projection of the bench. (Some people may not see the far probe as larger but almost all people see the far edge of the bench as larger than the near edge.) Experiments compared the occurrence of the size illusion in perspective and orthographic projections as a function of the separation of the near and far probes and as a function of the slant of the bench. For both perspective and orthographic projections, the illusion follows a similar course. The magnitude of the size illusion increased with the distance (measured by the number of contours separating the probes,e.g. 4 in Figures 4 and 5) between the near and far probes (Figure 6) and with the slant of the bench (Figure 7). The functions describing the size illusion for the perspective and orthographic projections were remarkably similar. The only difference is that the magnitude of the size illusion was greater by a small amount for the perspective projection. We have sought to explain the occurrence of a size illusion in an orthographic projection.

Figure 8 illustrates the size illusion in an orthographic projection of a rectangle. A size illusion occurs when a surface is seen in depth (top left figure) but not when it is seen in the plane (top right figure). The four lines at the corners of the bottom figure are all the same length. Subjects, however, consistently report that the line in Corner 4 is the longest and the line in Corner 1 is the shortest. We believe the size illusion can be used to identify the algorithm used by the visual system to perceive the tridimensional orientation of a pictured surface. The hypothesis

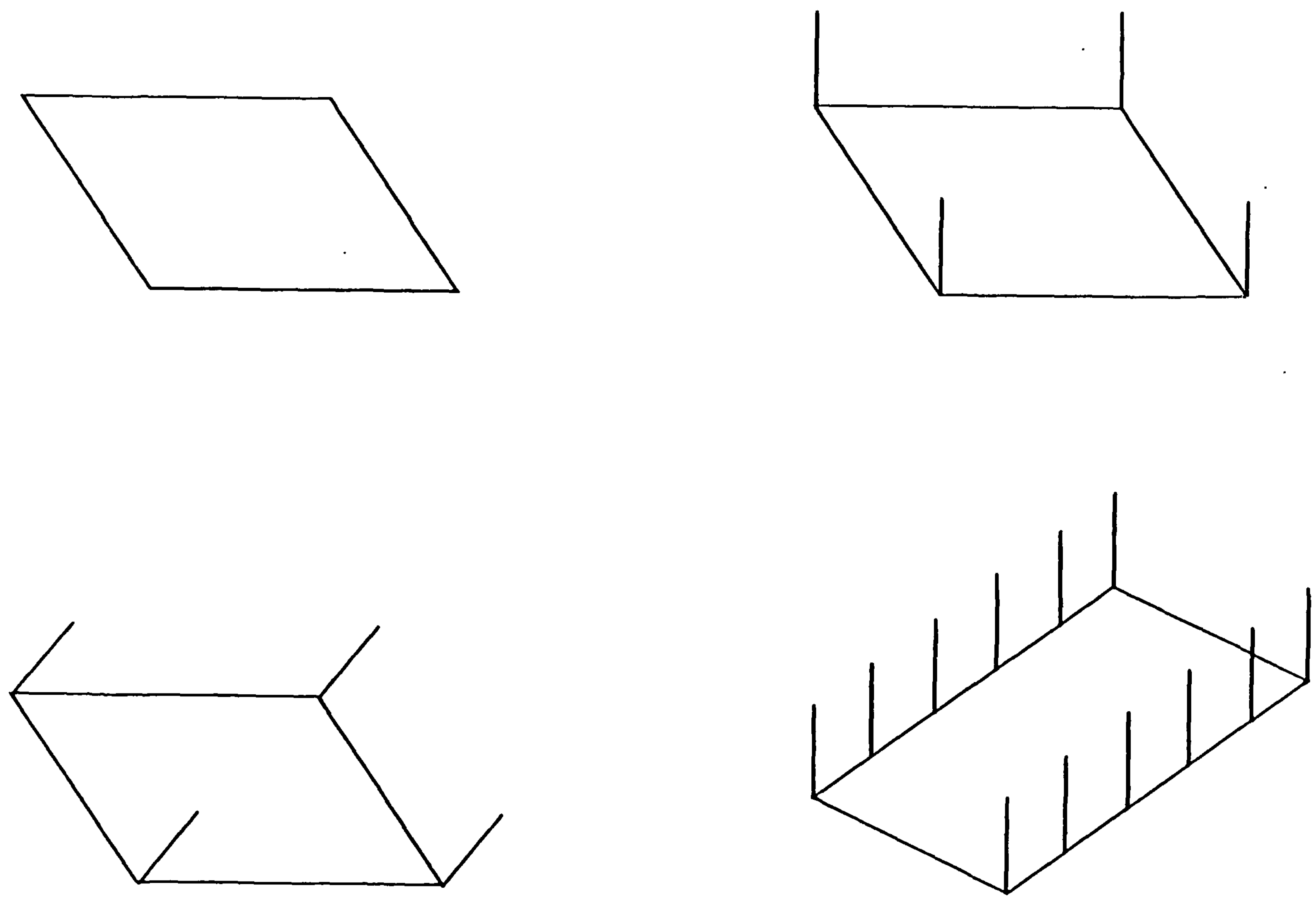


Figure 3

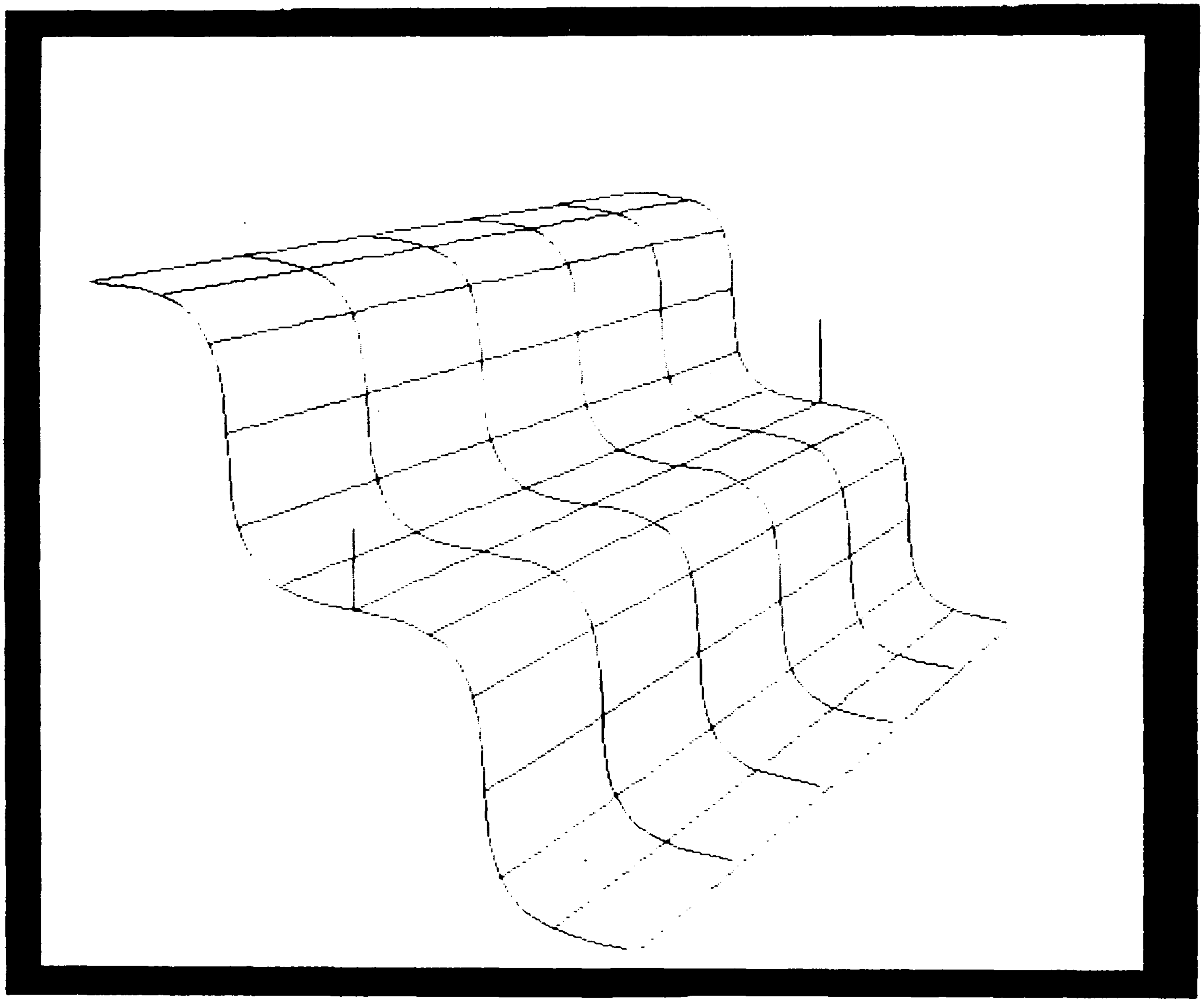


Figure 4

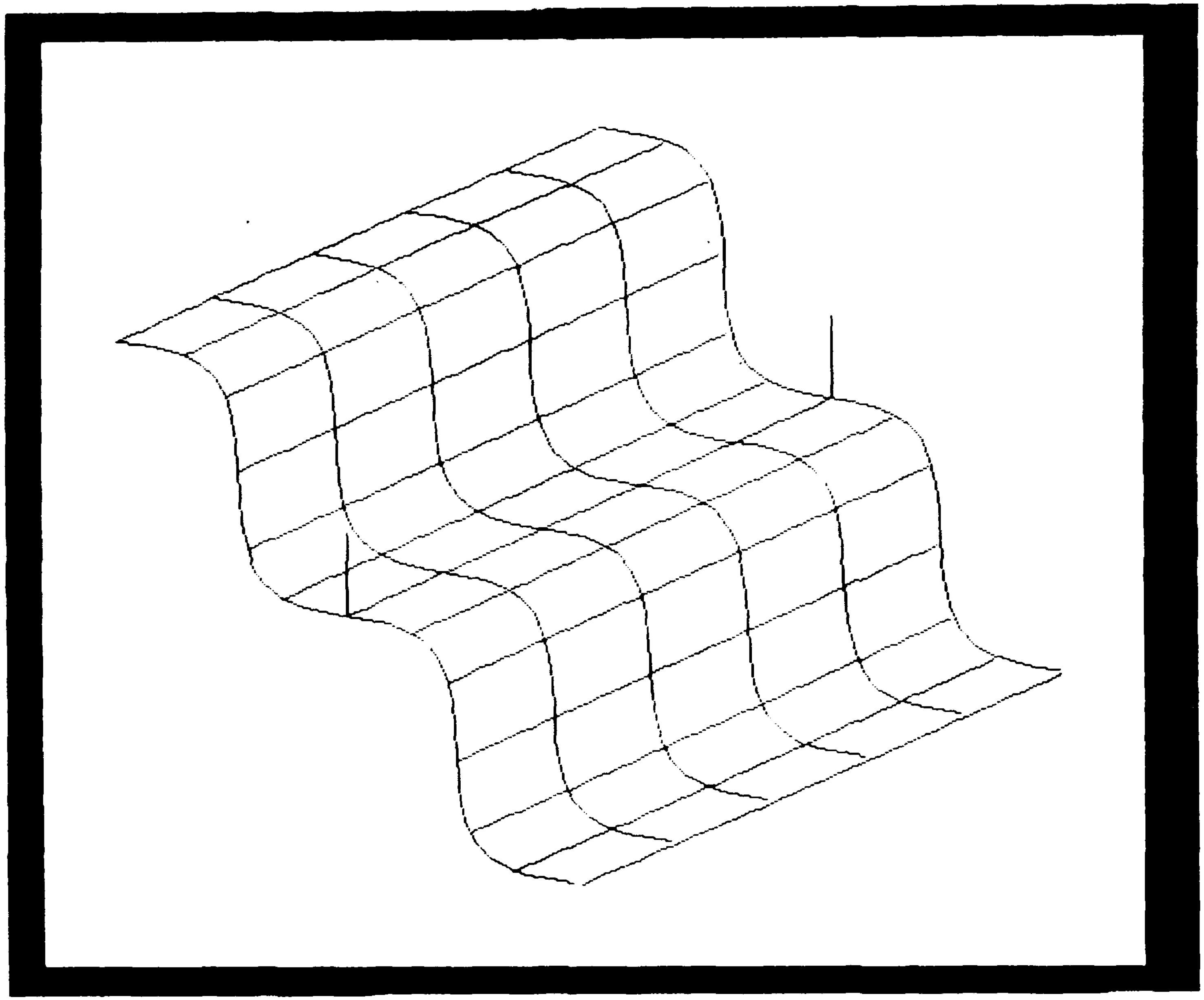
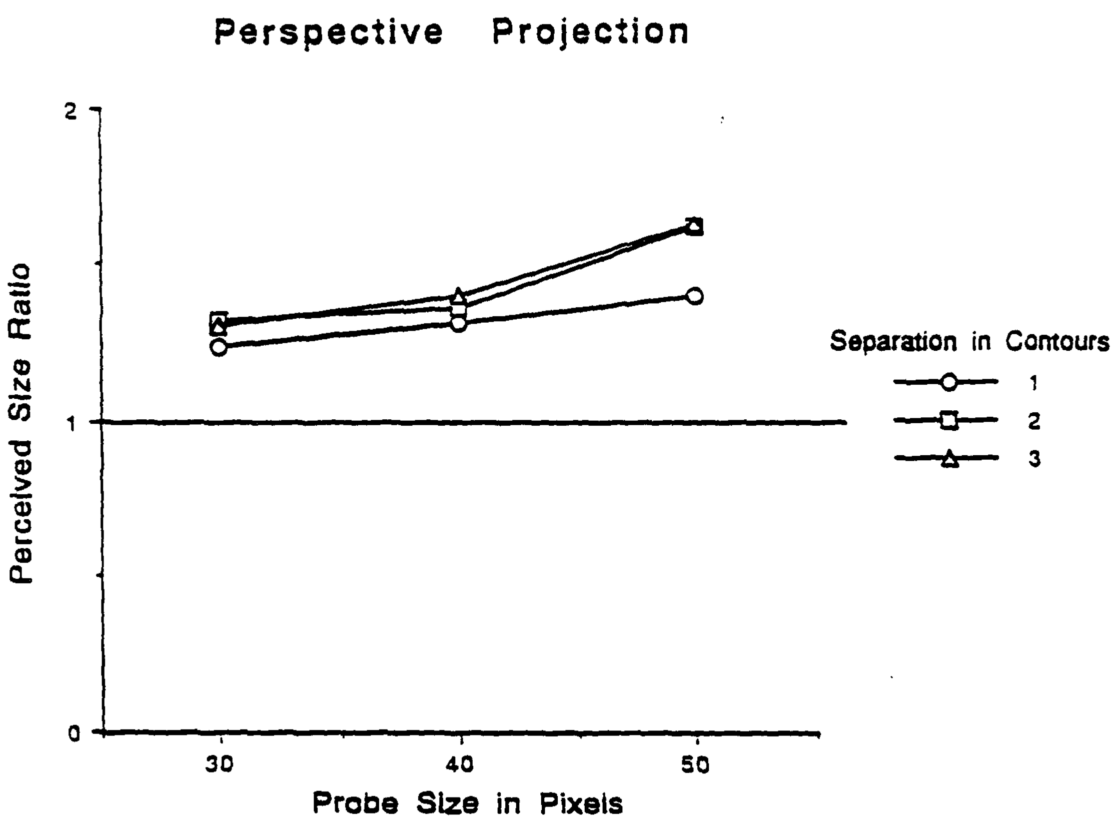


Figure 5


Perspective Projection
2
1
0
Perceived Size Ratio
30
40
50
Probe Size in Pixels
Separation in Contours
1
2
3

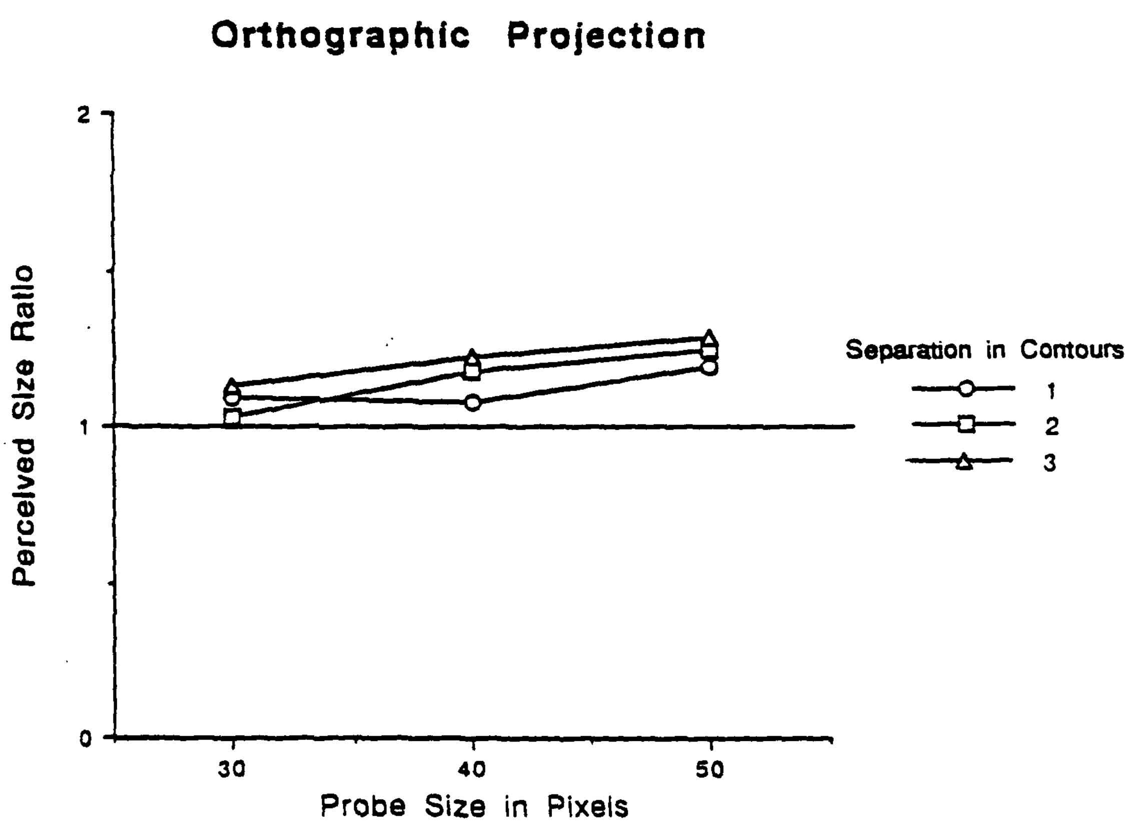
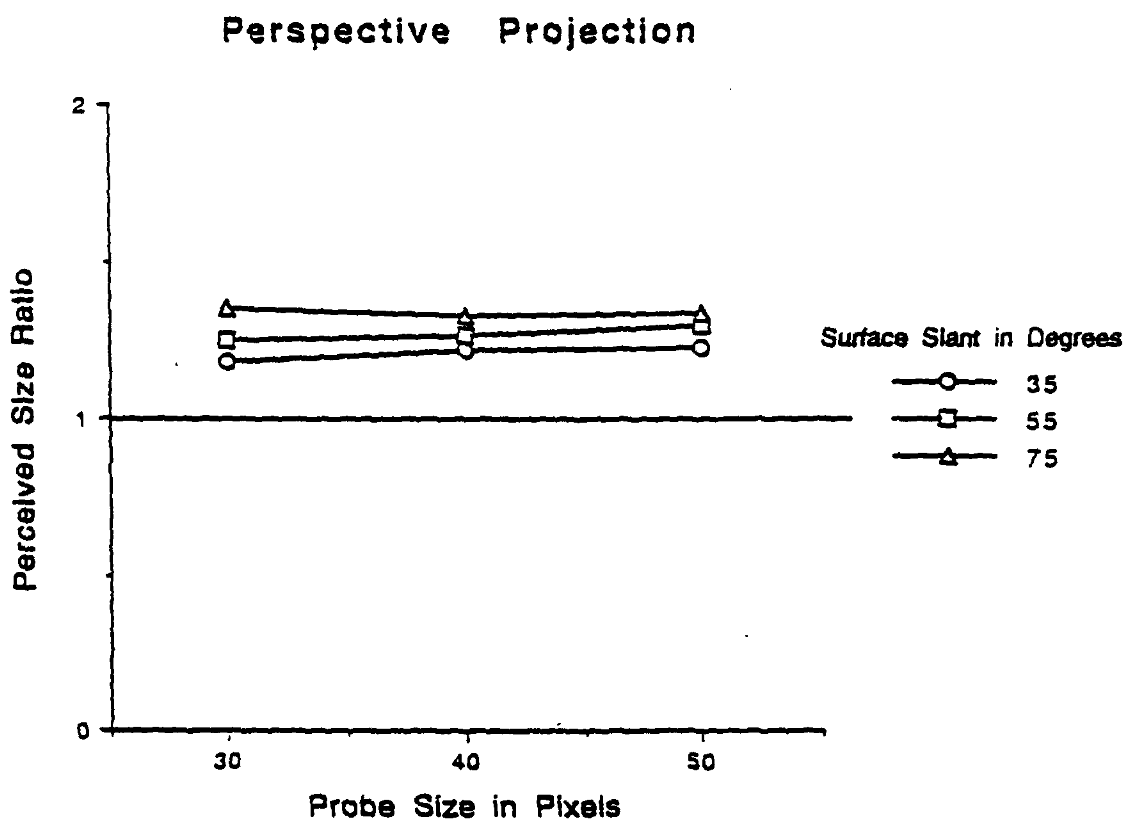

Orthographic Projection
2
1
0
Perceived Size Ratio
30
40
50
Probe Size in Pixels
Separation in Contours
1
2
3

Figure 6


Perspective Projection
2
1
0
Perceived Size Ratio
30
40
50
Probe Size in Pixels
Surface Slant in Degrees
35
55
75

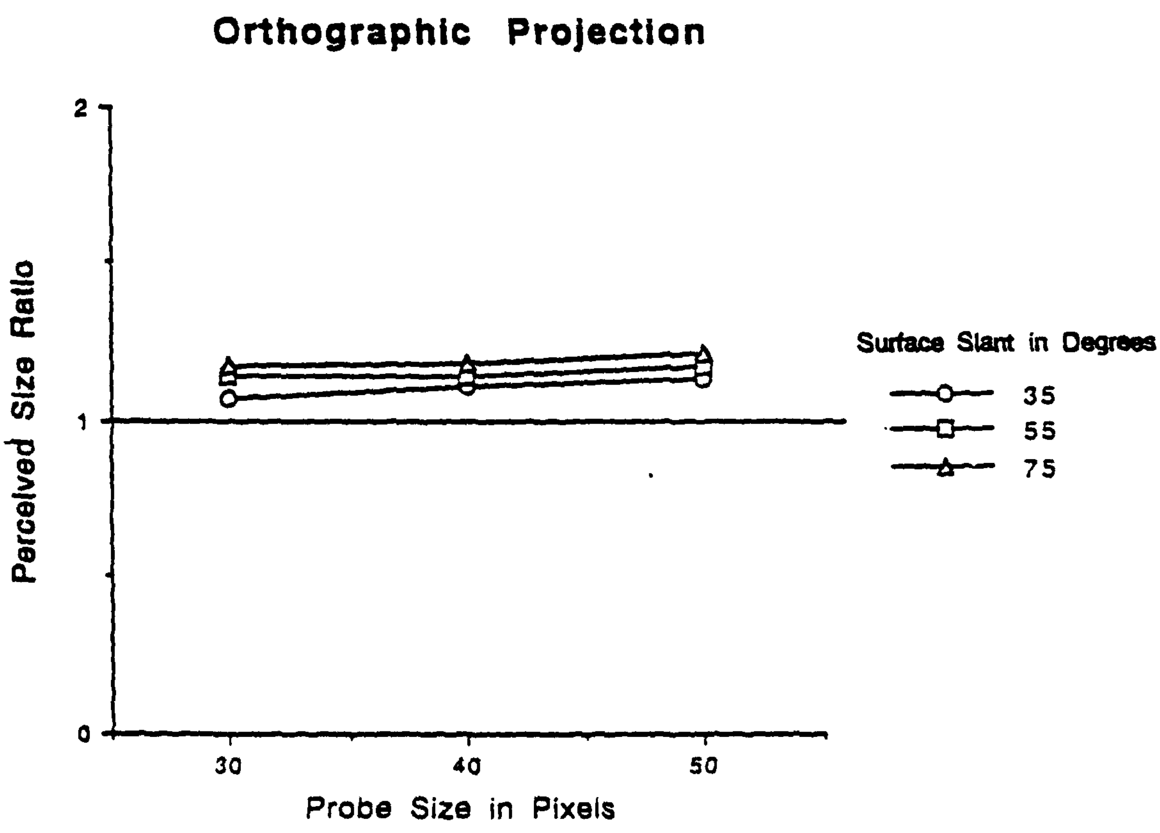

Orthographic Projection
2
1
0
Perceived Size Ratio
30
40
50
Probe Size in Pixels
Surface Slant in Degrees
35
55
75

Figure 7

that the visual system rotates a reference figure until a feature of its projection matches a feature of the pictorial display requires that the visual system render explicit an axis of rotation. The size illusion, we conjecture, can be used as a marker to determine the axis about which the reference figure is rotated.

Explanation of size illusion--One's first hypothesis is to ascribe the illusory perceptions of size to the size-distance relationship. In real space, objects which subtend the same visual angle are seen as larger when they are seen as further away. An illusory perception of size would, therefore, be produced by the normal mechanisms of size perception because of the perception of the greater distance of the line in Corner 4 than in Corner 1. If the illusion is due to the size-distance relationship, the magnitude of the size illusion should be a function of the distance of the observer. Our observations indicate that the size illusion is unaffected by the distance of observation. It is still possible, however, that a picture induces an apparent distance of observation that differs from the actual distance of observation and that remains constant with changes in observation distance. However, inverting Figure 8 shows that perceived distance can not be the sole factor. Now subjects see the near line (what was Corner 4 and is now Corner 1) longer and the far line (what was Corner 1 and is now Corner 4) shorter. The important point is that the longer line is now seen to be the line nearest to the observer in pictorial space and the shorter line is now seen to be the line furthest from the observer in pictorial space. Another possible factor is that the 'longer' line (Corner 4 in Figure 8 held upright) has outward pointing wings and the shorter 'line' (Corner 1 in the upright figure)has inward pointing wings as in the Muller-Lyer illusion. Again, this can not be be the complete explanation. Figure 3 (bottom right) shows the illusion occurs when the lines are not at the corners.

Why is there an illusory perception of size? Every picture can be seen in two ways. It can be seen to varying extent as what it physically is, a 2D image, and as what it represents, a 3D scene. It is well established that the perception of space in pictures shows regression to the 2D planar image. The size illusion, we propose, is due to the regression of the coordinates of the lines in the representation of 3D space to their coordinates in the 2D image. The bottom figure in Figure 8 illustrates the orthographic projection of a surface with lines in the corners. What is the relationship of the coordinates of the tops of the lines in Corners 1 and 4 when the pattern is seen as a 2D image and when the pattern is seen as a 3D image? The y-coordinate measures the height of a point above the ground or reference plane. The top figures in Figure 9 illustrate the 2D and 3D coordinates of the lines in Corners 1 and 4 with the lines pointing upward. The right figure illustrates the top y-coordinates of the lines in the 2D image. The lines are at the corners of the reference rectangle and lie in the plane of the figure. Measured from the base line in Figure 9, the top y-coordinate of the line in Corner 4 is 97 mm and of the line in Corner 1 is 48 mm. According to our model, the visual system slants the reference rectangle around a horizontal axis until the right angle projects into the foreshortened angle of the projected image. One should think of the lines as connected to the reference rectangle by flexible hinges and as the rectangle rotates the lines assume a perpendicular orientation to the surface in 3D space. Assume that the surface is slanted away from the observer. What happens to their top y-coordinates? The surface is slanted floorwise so that the top y-coordinate of the line in Corner 4 becomes less and the top y-coordinate of the line in Corner 1 becomes greater. The left figure illustrates the top y-coordinates of the lines in the 3D representation after rotation. Their heights above the base line in Figure 9 are 91 mm and 54 mm, respectively. Regression to the 2D coordinates lengthens the line in Corner 4 and shortens the line in Corner 1. What happens when the figures are inverted?

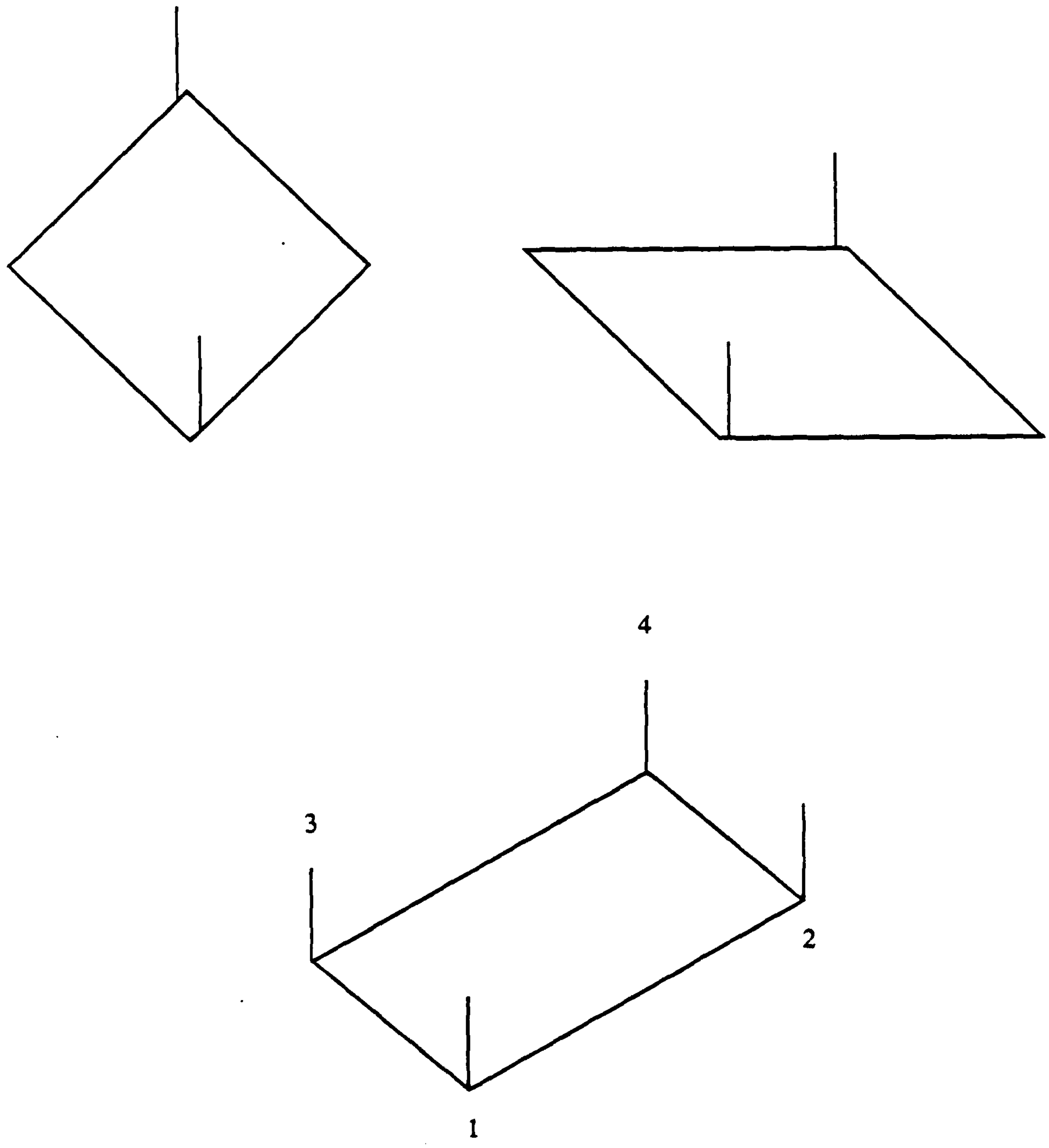


Figure 8

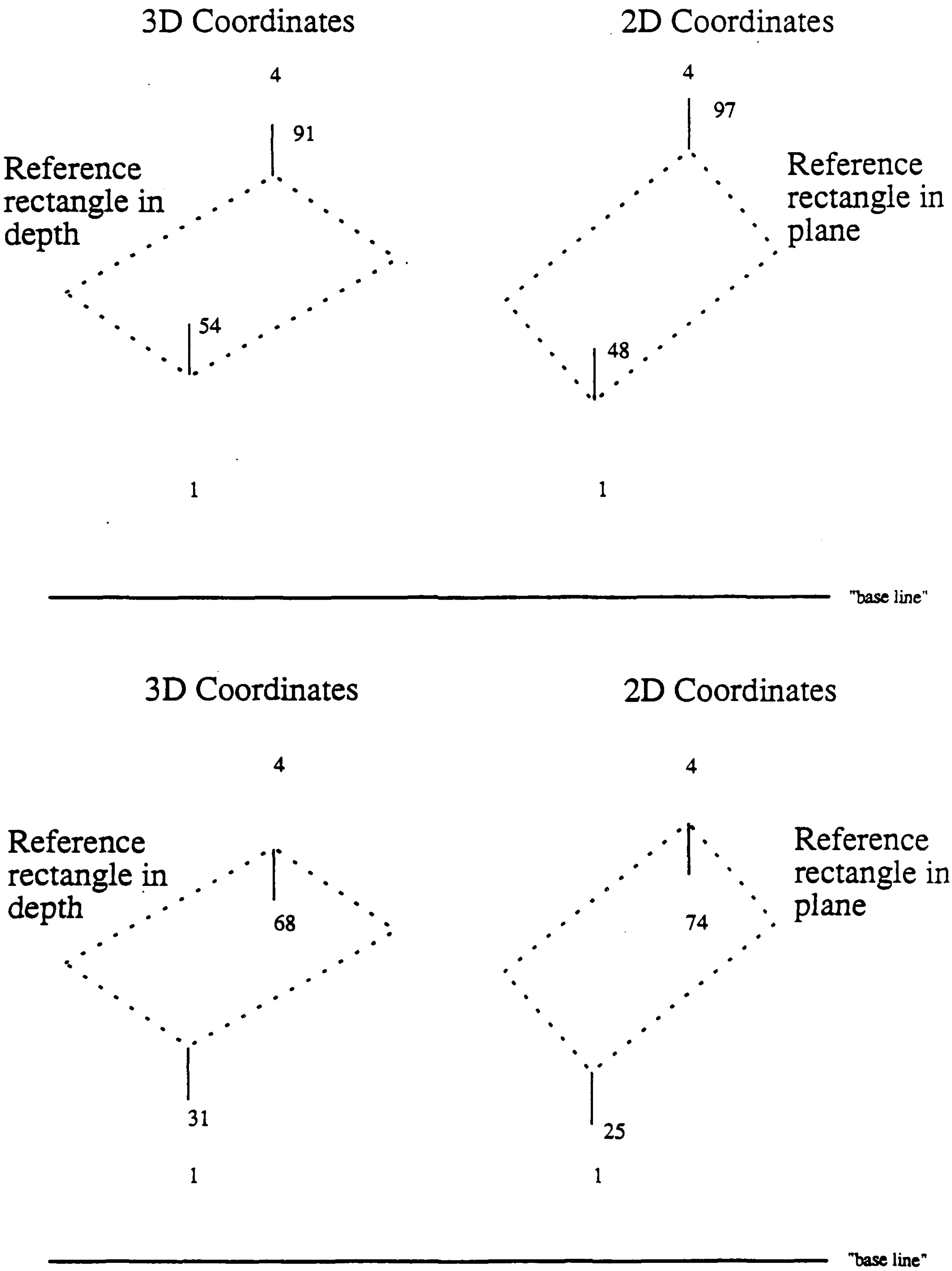

3D Coordinates
2D Coordinates
4
91
4
97
Reference rectangle in depth
Reference rectangle in plane
54
48
1
1
"base line"
3D Coordinates
2D Coordinates
4
4
Reference rectangle in depth
Reference rectangle in plane
68
74
31
25
1
1
"base line"

Figure 9

The bottom figures in Figure 9 show the top figures rotated 180 degrees. Now when the reference rectangle is slanted floorwise the bottom y-coordinate of the line in Corner 1 becomes less and the bottom y-coordinate of the line in Corner 4 becomes greater. Regression of the bottom y-coordinates to their 2D values now lengthens the line in Corner 1 and shortens the line in Corner 4. When the lines point ceilingwise and the surface is slanted away from the observer, regression causes lines in front of the axis of slant to look shorter, and lines behind the axis of the slant to look longer. When the lines point floorwise and the surface is slanted away from the observer, regression causes lines in front of the axis of slant to look longer and lines behind the axis of slant to look shorter. The greater the distance between the lines and the axis about which the reference rectangle is slanted, the more their 3D coordinates differ from their 2D coordinates. The change in the y-coordinate is equal to the distance of the line from the axis of slant times the sine of the slant angle.

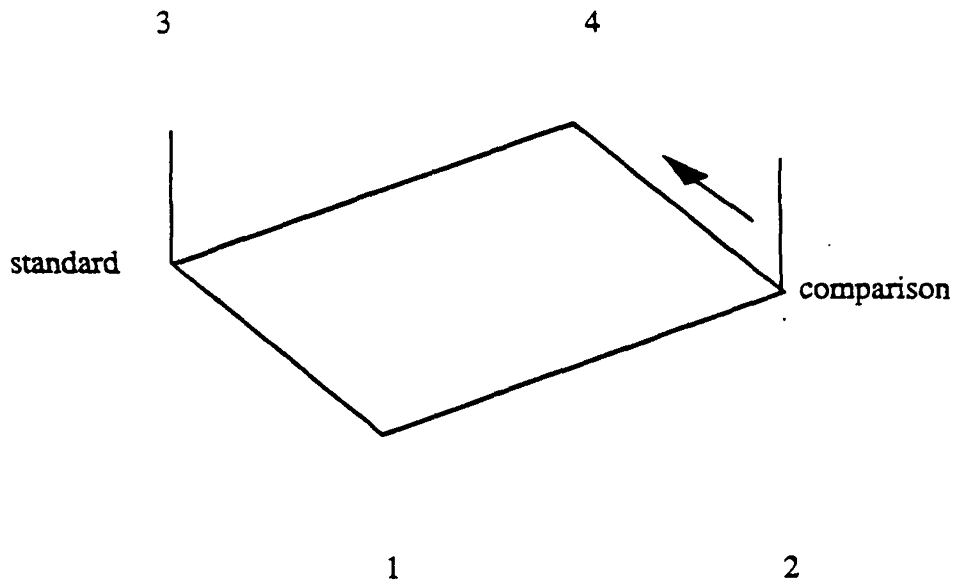
In an orthographic projection it is always possible to reverse the near and far edges. When this occurs in Figure 8, for example, the line in Corner 4 continues to be seen as the longest and the line in Corner 1 continues to be seen as the shortest. Reversing the near and far edges is equivalent to slanting a surface by the same amount toward the observer rather than away from the observer. Thus, regression of the coordinates of the lines in the representation of 3D space to their coordinates in the 2D image would lengthen and shorten the lines exactly in the same way as when the surface is slanted away form the observer. The only difference is in the formulation of our rule. Since near and far in the picture are reversed, our rule needs to be appropriately altered. When the lines point ceilingwise and the surface is slanted toward the observer, regression causes lines in front of the axis of slant to look longer, and lines behind the axis of the slant to look shorter. When the lines point floorwise and the surface is slanted toward the observer, regression causes lines in front of the axis of slant to look shorter and lines behind the axis of slant to look longer.

What happens to lines on the axis about which the reference rectangle is slanted? The 2D and 3D y-coordinates of lines on the axis of slant are the same. According to our hypothesis no illusion should then occur and the lines should be seen equal in length.

6.3 Experiments

We have conducted experiments to test whether lines located on the predicted axis of rotation will be seen equal in size. In one experiment, the lines were vertical. The presumption to see the lines as normals to the 3D surface makes the predicted axis of rotation horizontal. The stimuli were presented upright and inverted for 1 second. The method of constant stimuli was used. Subjects judged whether the comparison line was longer or shorter than the standard. The top figure in Figure 10 shows the standard and comparison lines with the lines pointing upward. The standard line was in Corner 3 and the comparison lines were located at 3 mm intervals from Corner 2 to midway between Corners 2 and 4. The bottom figure in Figure 10 shows the standard and comparison lines with the lines pointing downward. The standard line was in Corner 2 and the comparison lines were located at 3 mm intervals from Corner 3 to midway between Corners 3 and 1. Twenty subjects made three judgments each.

Figure 11 shows the proportions of times that the comparison line was judged longer than the standard with the lines pointing upward (top left) and the lines pointing downward (top right). The bottom left and bottom right figures in Figure 11 show the comparison lines at the predicted



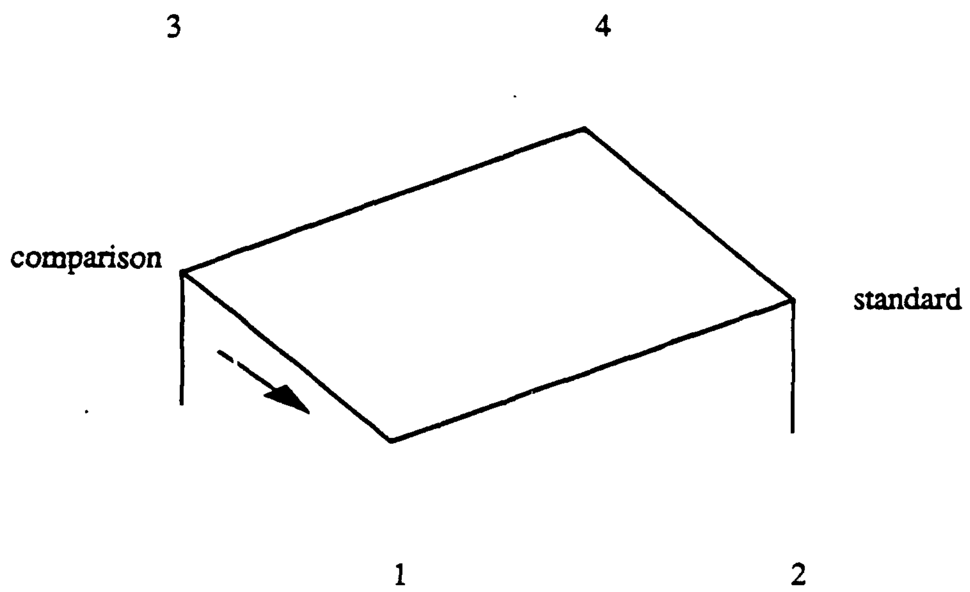


Figure 10

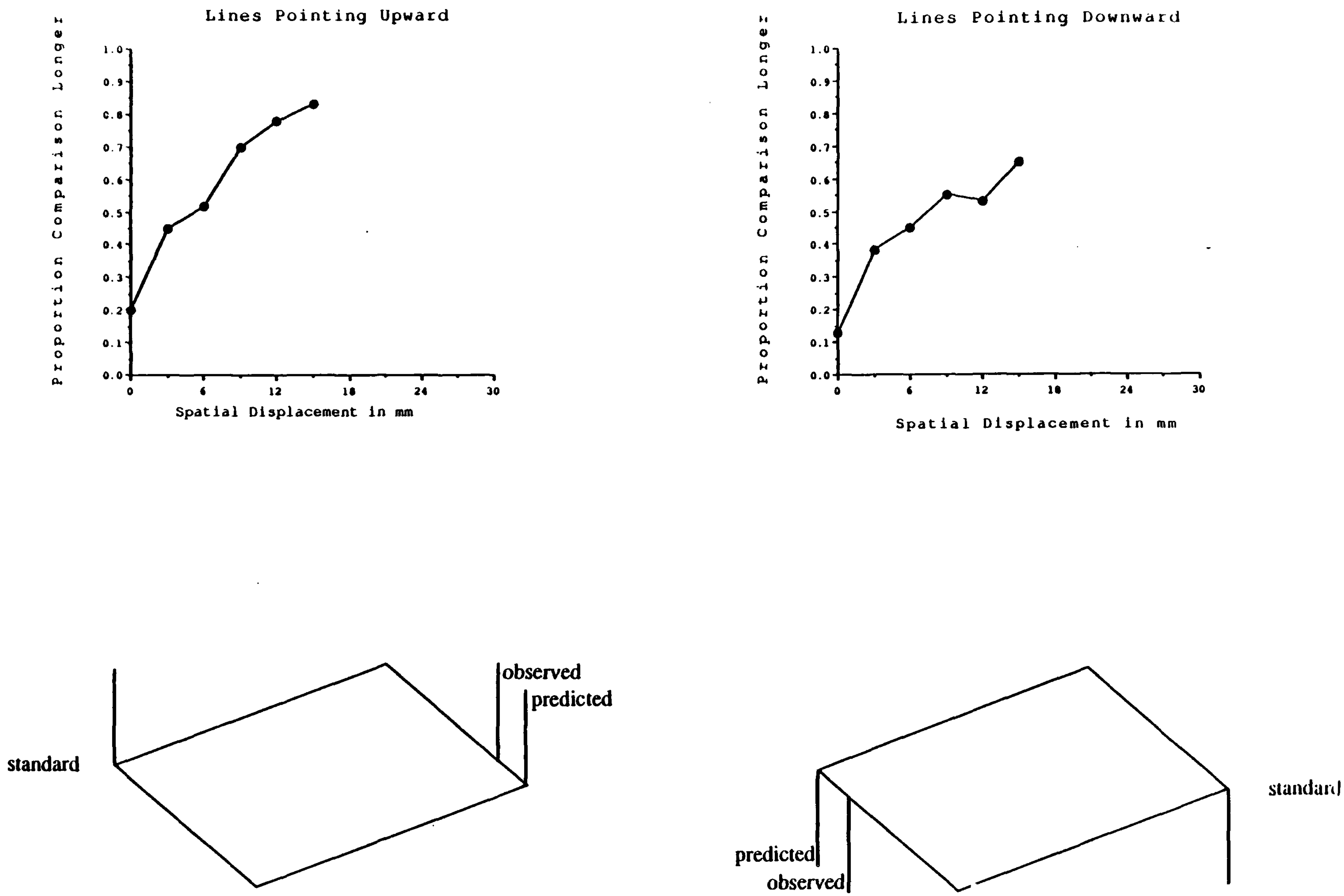

Lines Pointing Upward
Proportion Comparison Longer
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
6
12
18
24
30
Spatial Displacement in mm
Lines Pointing Downward
Proportion Comparison Longer
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
6
12
18
24
30
Spatial Displacement in mm
standard
observed
predicted
predicted
observed
standard

Figure 11

and observed equivalence points,i.e., the positions at which the proportion of longer judgments was .5. The predicted axis of rotation is 0 degrees and the axis of rotation derived from the experimental results is 6 degrees when the lines pointed upward and 8 degrees when the lines pointed downward. The constant errors are significant.

A second experiment was conducted in which the added lines were at 120 degrees and pointed upward. The comparison line was moved from Corner 2 to Corner 4 in 6 mm steps. Figure 12 presents the data. The predicted equivalence point is 19 mm; the obtained equivalence point is about 25 mm. The predicted axis of rotation is 28 degrees and the obtained axis of rotation is approximately 33 degrees. In a third experiment, the lines were slanted at 60 degrees from the horizontal. When the lines pointed upward, the standard was in Corner 3 and the comparison moved from Corner 2 to Corner 1 in steps of 6 mm (except for the last position which was 14 mm). When the lines pointed downward, the standard was in Corner 2 and the comparison moved from Corner 3 to Corner 4. Figure 13 presents the results. The predicted equivalence point is 50 mm from Corner 2 toward Corner 1 when the lines pointed upward and 50 mm from Corner 3 toward Corner 4 when the lines pointed downward. The obtained equivalence point is about 29 mm when the lines pointed upward and ranged from 9 mm to 43 mm when the lines pointed downward. The predicted axis of rotation is -30 degrees when the lines pointed upward and the axis of rotation derived from the data is -10 degrees.

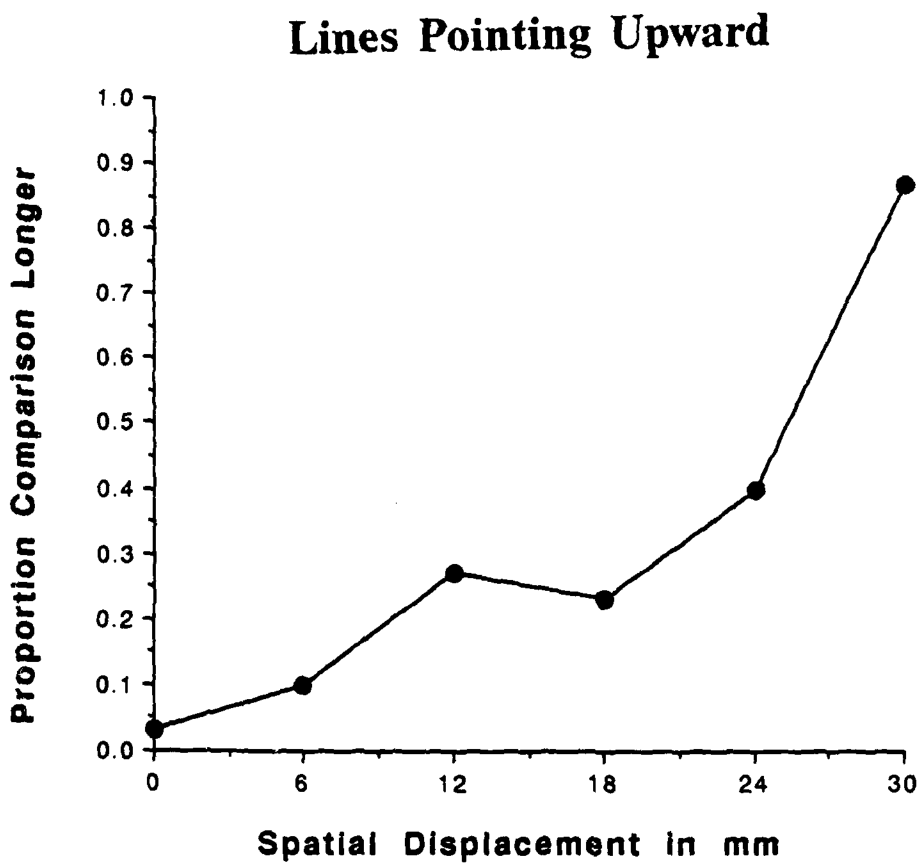
Except for the last experiment the axes of rotation are in agreement with the axes of rotation determined by the experiments reported in Section 5. The results provide provisional support the hypothesis that the visual system renders the axis of rotation explicit and that the visual system encodes the tridimensional of a surface by rotating a reference figure about the axis of rotation.

7. Illusory Perceptions of Size in Orthographic Projections

The top figures in Figure 14, modeled after Shepard (1981), illustrate a well known illusion. The length of the horizontal edge of the top right figure is the same as the length of the edge seen in depth in the top left figure. However, the length of the edge seen in depth in the left figure is seen as much longer. The illusion is qualitatively consistent with the hypothesis that the visual system in seeing the 3D figure in pictorial space carries out an inverse orthographic projection. It is not known, however, whether the magnitude of the illusion is quantitatively consistent with an orthographic projection.

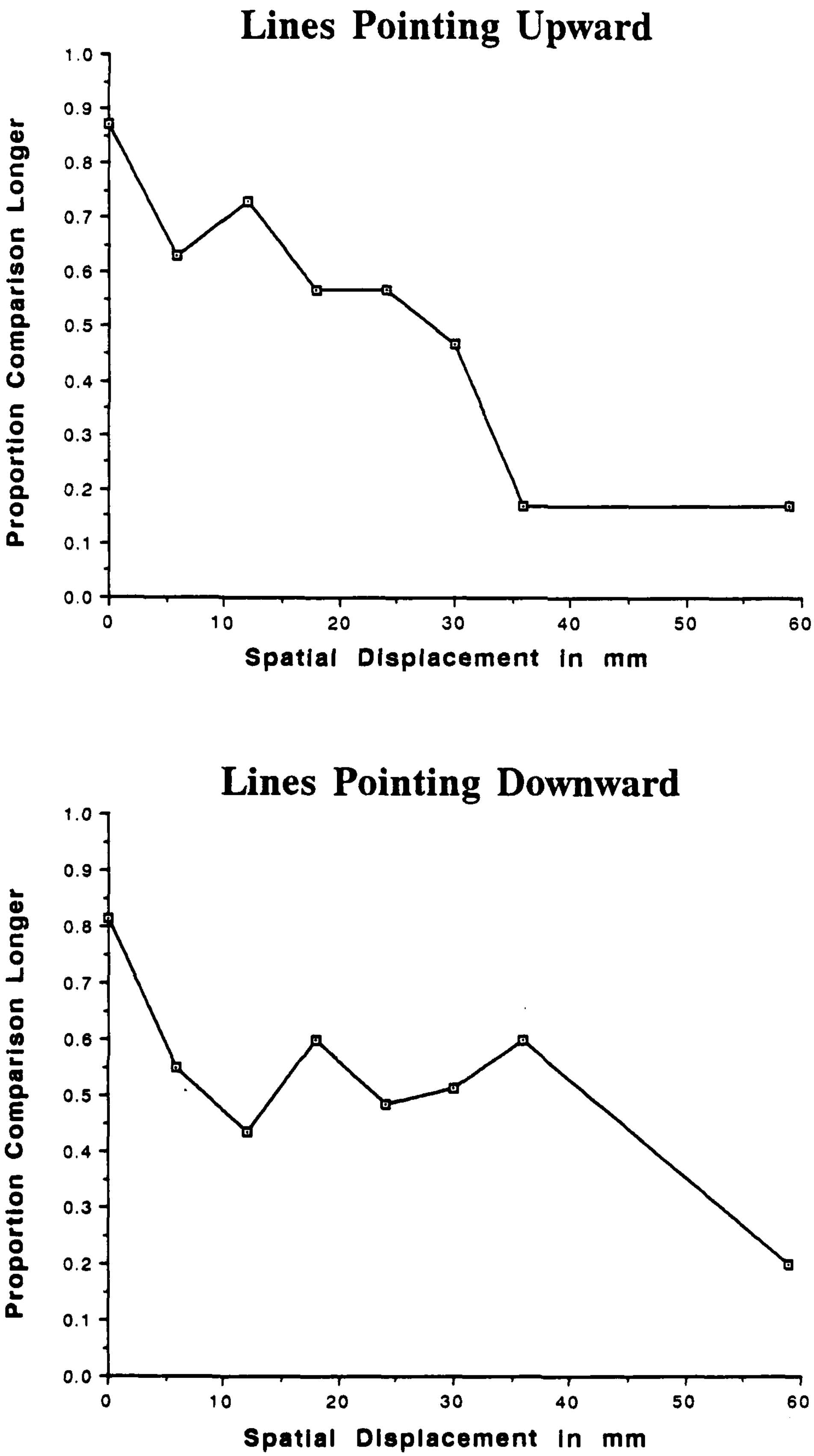
7.1 Experiments

An experiment tested whether the magnitude of the illusion is predicted by an orthographic projection. The bottom row in Figure 14 illustrates the stimuli used in the experiment. Subjects were asked to make size and orientation judgments of the top face of 11 orthographic projections of a box. The box was slanted back from the frontal plane about the bottom front horizontal edge and then rotated to the left about a vertical line through the bottom front left vertex by differing numbers of degrees. (The boxes shown in the bottom row of Figure 14 are rotated to the right and were not stimuli used in the experiment.) The length of the top edge of the modeled 3D box was always 100 pixels. The orthographic projection of this length in the picture plane differed for each stimulus ranging from 45 to 77 pixels depending on the slant and rotation of the box.



Lines Pointing Upward
Proportion Comparison Longer
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
6
12
18
24
30
Spatial Displacement in mm

Figure 12



Lines Pointing Upward
Proportion Comparison Longer
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
10
20
30
40
50
60
Spatial Displacement in mm
Lines Pointing Downward
Proportion Comparison Longer
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
10
20
30
40
50
60
Spatial Displacement in mm

Figure 13

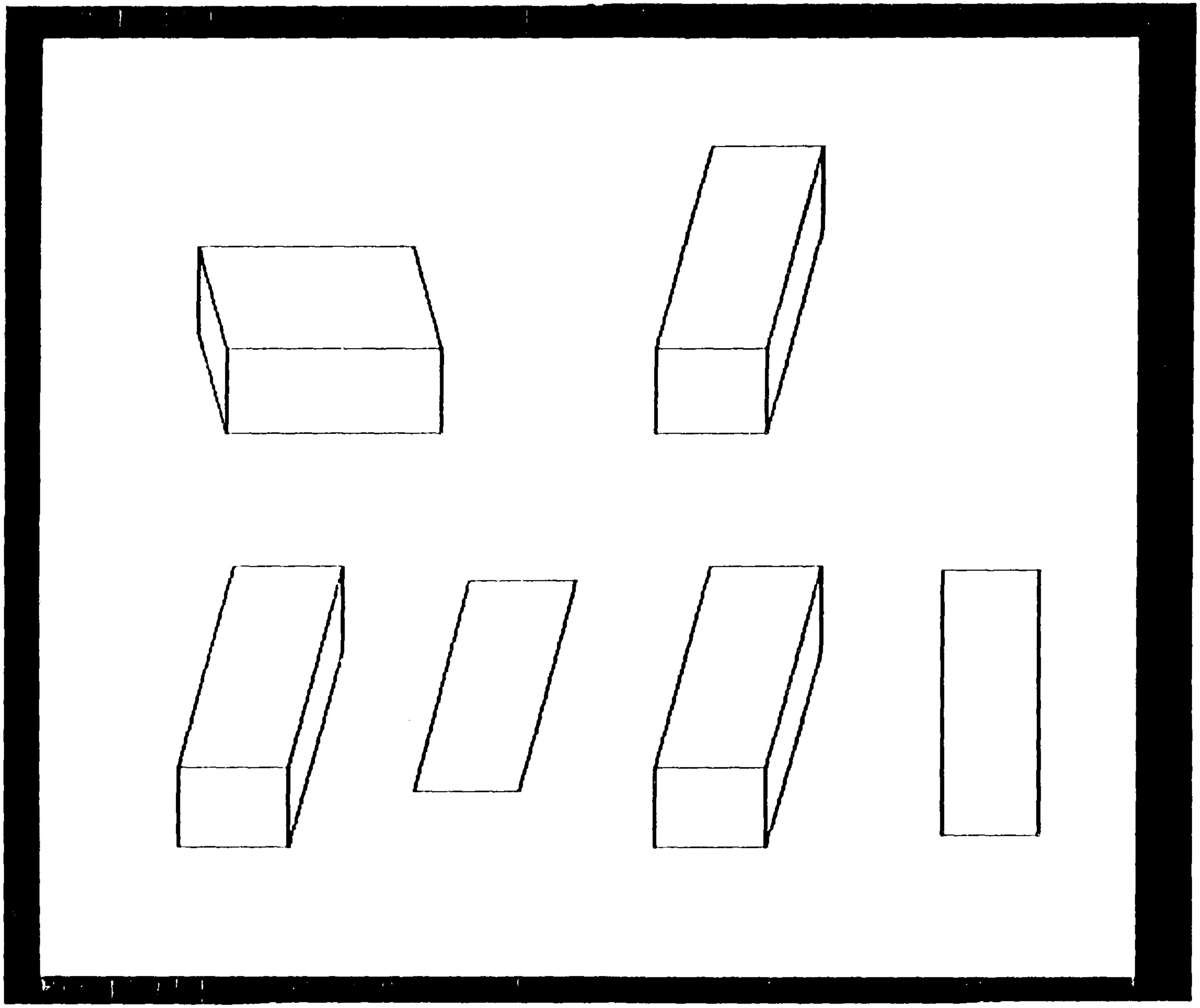


Figure 14

Nine subjects served in the experiment. A subject was first asked to match the size of a rectangle in the picture plane (bottom right) to the top surface of the box. Clicking the buttons of a mouse allowed subjects to increase or decrease both the width and length of the rectangle in 5 pixel increments. The instructions emphasized that it was especially important to carefully equate the perceived length of the rectangle to the perceived length of the top surface of the box. A subjects was then asked to judge the spatial orientation of the box by adjusting the slant and rotation of a rectangle to match the perceived spatial orientation of the box. Clicking the buttons of a mouse allowed subjects to increase or decrease both the slant and rotation of the rectangle in 5 degree increments. As subjects clicked, the computer immediately plotted the orthographic projections of the rectangle on the monitor screen. Subjects were given practice until they became proficient at clicking the mouse buttons quickly. When the mouse buttons were clicked quickly, one had the impression of the rectangle turning in space as in an animated movie. Each of the box stimuli was shown four times.

Table 6 presents the subjects' mean length judgments and the predicted length judgments from an orthographic projection. The predicted length judgments are a function of the perceived slants and rotations of the boxes. Although not shown in the table, subjects judgments of the perceived rotations of the boxes were accurate, i.e., they are very close to the actual rotations. Subjects, however, consistently underestimated the slants of the boxes, i.e., the mean perceived slants were consistently less than the actual slants of the boxes. Subjects' mean length judgments of the top surface were also less than the 100 pixel edge length of the 3D modeled box. Since subjects consistently underestimated the perceived slants of the boxes, the perceived length of the top surface would be expected to be less. Table 6 shows the predicted length judgments based upon subjects' mean slant and rotation judgments. A comparison of the obtained mean length

Table 6

Length Judgments

Stimuli	Mean Length Judgment	SD	Predicted Length Judgment
1	89.6	5.1	85.9
2	90.4	6.3	86.2
3	90.6	6.9	86.5
4	89.3	6.1	88.8
5	84.0	10.6	80.8
6	85.6	10.2	80.3
7	85.0	10.5	79.5
8	85.0	9.3	84.9
9	85.1	9.5	87.3
10	81.0	12.9	81.8
11	86.1	7.2	97.4

judgments and the predicted length judgments show that they are very similar except for stimulus 11. In a control experiment, each subject used a tilt board to judge the slant and rotation of the boxes. The slant and rotation judgments were similar to those obtained using the mouse.

The results are consistent with the hypothesis that the visual system carries out an inverse orthographic projection.

References

Attneave, F. (1972) Representation of physical space. In A.W. Melton & E. Martin (Eds.), *Cognitive processes in human memory*. New York: Wiley.

Attneave, F. & Frost, R. (1969) The determination of perceived tridimensional orientation by minimum criteria. *Perception & Psychophysics*, **6**, 391-396.

Beck, J., Tanaka J., & Goodwin, W. (1990) Perceptions of size in orthographic and perspective projections (in preparation).

Barnard, S.T. & Pentland, A.P. (1983) Three dimensional shape from line drawings. *Proceedings of Eighth International Joint Conference on Artificial Intelligence*, 1062-1064.

Koffka K., (1935) *Principles of Gestalt psychology*. New York: Harcourt Brace.

Mach, E. (1959) *The analysis of sensations*. New York: Dover.

Perkins, D.N. (1972) Visual discrimination between rectangular and nonrectangular parallelopipeds. *Perception & Psychophysics*, **12**, 396-400.

Perkins, D.N. (1973) Compensating for distortion in viewing pictures obliquely. *Perception and Psychophysics*, **14**, 13-18.

Perkins, D.N. (1983) Why the human perceiver is a bad machine. In J. Beck, B. Hope, & A. Rosenfeld (Eds.) *Human and Machine Vision*. New York: Academic Press.

Perkins, D.N. & Cooper, R.G., Jr. (1980) How the eye makes up what light leaves out. In M.A. Hagen (Ed.), *The perception of pictures*. New York: Academic Press.

Pirenne, M.H. (1970) *Optics, painting, and photography*. Cambridge: Cambridge University Press.

Shepard, R.N. (1981) Psychophysical Complementarity. In M. Kubovy & J.R. Pomerantz (Eds.), *Perceptual Organization*, New Jersey: Lawrence Erlbaum Associates.

Shepard, R. N. & Metzler, J. (1971) Mental rotation of three-dimensional objects. *Science*, **171**, 701-703.

Smith, O.W. (1958) Comparison of apparent depth in a photograph viewed from two distances. *Perceptual and Motor Skills*, **8**, 79-81.

Stevens, K.A. (1981) The visual interpretation of surface contours. *Artificial Intelligence*, **17**, 47-73.

Stevens, K.A. (1983) Surface tilt (the direction of slant): A neglected psychophysical variable.

Perception & Psychophysics, 33, 241-250.

Stevens, K.A. (1986) Inferring shape from contours across surfaces. In A.P. Pentland (Ed.), *From Pixels to Predicates*. Norwood, NJ: Ablex.

Wallach, H. & Slaughter, V. (1988) Viewing direction and pictorial representation. *Perception & Psychophysics*, 43, 79-82.

Publications and Meetings

Beck, J., Tanaka, J., & Goodwin, W. Perceptions of size in orthographic and perspective projections (in preparation).

Goodwin, W., & Beck, J. 3D interpretations of orthographic projections (in preparation).

Tanaka, J. (1987) Judging interobject distance in a pictorial scene. Unpublished First Year Research Project, University of Oregon.

Tanaka, J., Goodwin, W., & Beck, J. (1989) Perceptions of size in orthographic and perspective projections. Presented at the meetings of the Psychonomic Society, Atlanta, Georgia.

Personnel

John Bainbridge, Graduate Student
William Goodwin, Mathematician and Programmer
Steve Jones, Graduate Student
Patrick Jennings, Undergraduate Student
Tracey Redaway, Undergraduate Student